OULUN YLIOPISTO

Ermela Bufi

Finnish Fourth Graders' Number Sense and Related Misconceptions in Mathematics Learning

- A study of pupils' performance in judging the reasonableness of computational results and their reasoning strategies -

Master's Thesis

KASVATUSTIETEIDEN TIEDEKUNTA
Intercultural Teacher Education

23.10.2019

Faculty of Education
Finnish Fourth Graders' Number Sense and Related Misconceptions in Mathematics Learning: A study of pupils' performance in judging the reasonableness of computational results and their reasoning strategies
(Ermela Bufi)
Master's thesis, 76 pages, 1 appendix
October, 2019

Number sense, broadly defined as a general understanding of numbers and mathematical operations, is developed through instruction from an innate primitive ability to grasp quantity changes, into complex skills to engage with complex algorithms. The paramount importance of number sense in mathematics learning has been emphasized worldwide in mathematics education research and curricula setting since the 1980s. One of the main identified components of number sense is the learner's ability to judge the reasonableness of computational results. This ability is also emphasized in the recent Finnish National Core Curriculum for Basic Education 2014.

The aim of this research was to investigate Finnish fourth graders' number sense and related misconceptions they reveal in their mathematics learning. The study measured the performance of 90 fourth graders from a school in Northern Finland in judging the reasonableness of computational results and analyzed the solution strategies pupils employed when answering the questions. A web-based two-tier diagnostic test was used for such a purpose. The test was based on instruments used in similar research internationally and it was adapted in line with the curriculum and learning materials used locally.

Results revealed that the study participants perform less well in identifying reasonable and meaningful answers to mathematical problems, compared to how they perform in typical pen-and-paper mathematical calculations. The average correct answer rate for all the ten questions of the test was 57%. These findings are in line with prior research, which has found pupils' number sense ability to lag behind their mechanical computational skills. In 28% of the cases sampled pupils revealed various mathematics misconceptions due to incorrect modelling, non-mathematical prototypes, overgeneralizing of knowledge, or challenges in linking mathematical process-object linking.

This thesis provides an in-depth analysis of pupils' answers and their captured reasoning, drawing on the theoretical concepts of number sense and misconceptions in mathematics, as well as findings reported in related prior research. Several implications for developers of learning materials and teachers are discussed and a list of recommendations for further research is provided. The study adds to the international mathematics education research on number sense and contributes to the Finnish national discussion on improving curricula and instruction for higher mathematics proficiency among all learners.

Keywords: number sense; reasonableness of computational results; misconceptions; mathematics performance; fourth grade

Acknowledgements

To my daughters, who inspire daily

a most genuine interest in how children feel and think

I would like to extend my profound gratitude to the principal and the fourth grade teachers and pupils, who enabled the data collection for this research. Thank you for your collaboration throughout the whole process. The dedication of your school towards meaningful learning and embetterment has inspired me.

My sincere appreciation goes to my thesis supervisor Katri Jokikokko, as well as to Sonja Lutovac and Taiju Marttila for all their comments on earlier drafts. Your thoughtful remarks have significantly improved this work.

Last, I feel indebted to my family and friends. Your love and support provides meaning to everything I do. Thank you for being present.

With gratitude,

Ermela Bufi

October, 2019

Contents

1 Introduction

Number sense, broadly understood as a general understanding of numbers and mathematical operations (McIntosh et al., 1997) forms the very foundation for learning and understanding in mathematics. While in its most primary form it is an innate ability of the human and few other species, number sense needs to be developed through both formal and informal instruction in order for it to enable further conceptual understanding and skills in the mathematics domain (Berch, 2005). Underdeveloped number sense may lead to dyscalculia and other learning difficulties in mathematics (Mazzocco et al., 2011) or innumeracy, defined as a state where the learner can perform mathematical operations mechanically, but does not properly comprehend them and as a result draws false conclusions (Paulus, 1988). Due to the importance that number sense has for meaningful learning in mathematics, the concept has received much attention since the 1980s among cognitive psychology and mathematics education research communities (Dehaene, 2011; Kalchman, Moss, & Case, 2001), as well as in mathematics curricula setting circles (NCTM, 2000).

While number sense exhibits itself in a multifaceted array of skills (Berch, 2005), one of its main components is the ability to judge the reasonableness of a computational result (Kalchman, Moss, & Case, 2001; Yang, 2019). This ability refers to determining whether an obtained answer to a mathematical problem is acceptable, judging by how such a result compares with the expectations that the learner had built about the result, based on his or her prior knowledge and understanding, or judging by whether such a result is meaningful in a real-world scenario (Alajmi & Reys, 2010). Studies that have measured the performance of basic education pupils in different components of number sense have reported that the area where pupils perform the lowest is that of recognizing reasonable answers to mathematical problems (Yang, Li & Lin, 2008; Mohamed & Johnny, 2010).

The National Council of Teachers of Mathematics (NCTM)—the world's largest mathematics education organization founded in 1920 in the United States and Canada and operating nowadays as the most important actor in mathematics education standard setting and publishing worldwide—has emphasized in the *Principles and Standards for School Mathematics* (2000) that pupils should be able to employ a variety of strategies to judge the reasonableness of numerical computations and their results. While a similar objective was missing from the 2004 edition of the Finnish national curriculum for basic education, the most recent edition of 2014 lists among the central objectives of instruction in mathematics in grades 3-6 that the pupil should develop skills "in assessing whether the solution is reasonable and meaningful" (Finnish National Board of Education[1], 2014, p. 252).

Often, pupil performance in recognizing unreasonable answers and other areas of number sense is hindered by misconceptions that they hold. In general, misconceptions refer to constructs that learners develop to make sense of the world around them, when such constructs contradict accepted science and as a result lead to incorrectly constructed scientific concepts (Allen, 2014). In mathematics, misconceptions often steam from improper modelling of the problem, the use of non-mathematical prototypes, overgeneralizations of knowledge, and challenges in process-object linking (Ryan & Williams, 2007). Identifying learner's misconceptions and replacing them with well-constructed scientific concepts is the bases of meaningful learning, although a very challenging task for educators (Novak, 2002).

This study aims to assess fourth graders' number sense by measuring their performance in judging the reasonableness of computational results, as well as to identify some of the misconceptions pupils rely on when identifying reasonable answers to mathematical problems. The research questions addressed in the study are therefore:

(1) How do Finnish fourth graders perform in the number sense component of judging the reasonableness of computational results?

(2) What misconceptions do Finnish fourth graders show when trying to identify whether a given result is a reasonable answer to a mathematical problem?

An investigation of these questions contributes both to the international mathematics education research literature on number sense, as well as to the Finnish discussion on teacher training and curriculum development.

Although research on number sense is abundant, studies focusing specifically on measuring number sense and assessing the ability to judge the reasonableness of answers are quite limited. Moreover, as it is further discussed in section 2.4, this research is dominated only by a handful of researchers leading potentially to some extent of academic bias in the investigation of the phenomenon. While the pioneering research on measuring number sense was conducted in the United States, Australia, Taiwan, and Sweden (McIntosh et al., 1997), the overwhelming majority of subsequent research on the matter has been conducted in Asian countries, where the

[1] Nowadays: Finnish National Agency for Education

national mathematics curricula does not include the development of the ability to judge reasonableness of computational results in its instructional objectives (Alajmi & Reys, 2010; Yang, 2019). This study would be a welcomed addition to this current research landscape, as it addresses a few of its current limitations: it adds Finland, a Western, European country to the list of academic settings where number sense has been assessed recently; and it measures performance in judging the reasonableness of a computational result in a setting where the national curriculum has included the skill among its learning objectives for basic education mathematics. Prior to this study, number sense has been assessed for research purposes in Finland only in the pre-school age group (Aunio et al., 2004; Aunio et al., 2006).

The importance of high quality instruction for the development of number sense is well-established (Berch, 2005; Dehaene, 2011). Yet, research has shown that teachers have an inaccurate view of their pupils' level of number sense. As McIntosh et al. (1997, p. 6) state:

> "Although [teachers] value understanding, their instructional techniques and strategies do not necessarily support the development of number sense and most of them do not recognize this mismatch between their beliefs and practices."

More recent research supports such findings. For example, teachers emphasize computational operations to derive the exact answer over strategies that indicate whether a result could be an acceptable answer to a problem, since in their view, mathematics is an exact science and as such the exact answer derived through logical algorithms is the only reasonable answer (Alajmi & Reys, 2007). In Finland the new curriculum has added the ability to recognize reasonable and meaningful answers to mathematical problems, and the pre-service teacher training to some extent reflects such a change, but there might be a need for further teacher awareness and training both in the pre-service and in-service stage. This study may provide teachers with tools to assess their pupils' understanding of number and mathematical operations and it raises teachers' awareness of the effect of instruction on the development of number sense. As such, this research contributes to the on-going Finnish discussion on the need to improve mathematics learning, sparked by the national decline in PISA results in mathematics (Metsämuuronen & Tuohilampi, 2014).

This study recorded the performance judging the reasonableness of computational results and related misconceptions of 90 fourth-grade pupils (aged 10-11) in a school in Northern Finland. The group age was chosen based on the current developmental model of number sense as a function of age, which shows that by the age of 10 children typically have developed a full understanding of whole numbers, enabling them to solve problems requiring estimation and mental arithmetics (Kalchman, Moss & Case, 2001). A web-based two-tier test was used to measure the performance in recognizing reasonable answers (tier 1) and to identify the reasoning strategies that pupils employ when deriving their answers (tier 2). Pupils' answers from tier 2 were then classified based on whether the used strategy relied on a well-developed number sense (NS), on the application of a rule known from mathematics operations (R), a misconception about number magnitudes or the effect of operations on numbers (M), guessing (G), or other factors (O). The frequencies of the given categories were reported and a Chi-square test was performed to determine whether the observed distribution of the answers differed significantly from a random distribution explained solely by chance.

This thesis is organized as follows: the theoretical background elaborates on the main concepts of the research questions, namely those of number sense, ability to judge the reasonableness of computational results, and misconceptions, while concluding with a review of prior research in the field. Section 3 elaborates on the methods of collecting and analyzing the data used in this research, the findings from which are reported in the next section 4. A broader discussion of the results, what implications they have for curricula and teachers, as well as how they compare with prior research findings in the field is provided in the conclusion. The thesis ends by outlining some research reliability issues and providing recommendations for further research.

2 Theoretical background

This section provides the theoretical framework that is at the backbone of this study. The central concept is that of number sense. Section 2.1 elaborates on the definition of number sense, its importance in the development of mathematical knowledge and skills, and how it is developed from a primary intuition about numbers into the ability to understand and carry out complex algorithms. Section 2.2 focuses on one of the main components of number sense as defined by researchers and curriculum setters, namely the ability to judge the reasonableness of a computational result. The section elaborates both on its importance in mathematics learning, as well as on the challenges educators may face in seeing it as an integral part of mathematical skills. As one of the aims of this research is the identification of misconceptions that children have in mathematics learning, section 2.3 focuses exclusively on that, by elaborating, first, on general misconceptions and their role in learning, and subsequently, focusing on the nature of misconceptions in mathematics specifically. This theoretical framework is concluded with a review of the research that has focused on similar issues as this present study, starting with a brief description of each study and concluding with general summary of the current knowledge in the field, as well as a critical assessment of this knowledge.

2.1 Number sense

"Number sense" was created as a term by Tobias Dantzig in 1954 to refer to humans' primary form of numerical intuition:

> "Man, even in the lower stages of development, possess a faculty which, for want of a better name, I shall call *number sense*. This faculty permits him to recognize that something has changed in a small collection when, without his direct knowledge, an object has been removed or added to the collection" (qtd. in Dehaene, 1997, p. 5).

Number sense was coined as a term in mathematics education in the 1980s and has been widely used since then both in mathematics teaching and especially in mathematics education research. Gersten et al. (2005, p. 296) have correctly pointed out that "no two researchers have defined number sense in precisely the same fashion." Berch (2005, p. 333) would add that this definition challenge becomes even more multifaceted, as cognitive scientists and math educators define the concept of number sense in different ways. Although it is difficult to find one universal and precise definition for it, several definitions that share many similarities are offered. Howden (1989, p. 6) described it as a “special “feel” for numbers, an intuition about how they are related to each other [sic] and the world around them.” Other researchers have referred to it as a “well-organized conceptual framework of number information” (Bobis, 1996), or as a “general understanding of number and operations” (McIntosh et al., 1997, p. 3).

For Howden (1989, p. 6) pupils manifest number sense when they can relate numbers to their experiences and are able to create extensions to those experiences. She makes a comparison between how two different classrooms responded to her request “Tell me the first thing that comes to your mind when I say twenty-four”. In one class, pupils produced answers such as “two dimes and four pennies”, “two ten rods and four one cubes”, “two dozen eggs”, “my uncle became 24 years old on Sunday”, “I will be 24 years old in 19 years”, “the day before Christmas”, etc. In the other class, pupils traced in the air the number 24, found it in the classroom calendar page, or said that 24 appears on a digital watch every hour. Howden (1989) concludes that the first group is revealing a better developed number sense, as the spontaneous and diverse answers reflect a high ability to connect numbers’ mathematical meaning with children’s own everyday life situations.

Drawing on the relevant literature in the domains of cognitive development, mathematical cognition, and mathematical education, Berch (2005) comprises a list of all the features how number sense presumably manifests itself (figure 1). A good understanding of how number sense is defined and what it is comprised of is important, as it has implications on whether this ability is perceived as something that can be developed and taught, as well as on how it is measured.

Number sense is what enables children’s mathematical thinking and problem solving. The National Council of Teachers of Mathematics (NCTM)—the world’s largest mathematics education organization founded in 1920 in the United States and Canada and operating nowadays as the most important actor in mathematics education standard setting and publishing worldwide—listed a number of skills that number sense was composed of (NCTM, 1987). According to this report “children with good number sense (1) have well-understood number meanings, (2) have developed multiple relationships among numbers, (3) recognize the relative magnitude of numbers, and (4) know the relative effect of operating on numbers” (p. 37). McIntosh et al. (1997) state that when number sense develops, children can use their understanding of numbers “to make mathematical judgments and to develop useful and efficient strategies for managing numerical situations” (p.3).

1. A faculty permitting the recognition that something has changed in a small collection when, without direct knowledge, an object has been removed or added to the collection (Dantzig, 1954).
2. Elementary abilities or intuitions about numbers and arithmetic.
3. Ability to approximate or estimate.
4. Ability to make numerical magnitude comparisons.
5. Ability to decompose numbers naturally.
6. Ability to develop useful strategies to solve complex problems.
7. Ability to use the relationships among arithmetic operations to understand the base-10 number system.
8. Ability to use numbers and quantitative methods to communicate, process, and interpret information.
9. Awareness of various levels of accuracy and sensitivity for the reasonableness of calculations.
10. A desire to make sense of numerical situations by looking for links between new information and previously acquired knowledge.
11. Possessing knowledge of the effects of operations on numbers.
12. Possessing fluency and flexibility with numbers.
13. Can understand number meanings.
14. Can understand multiple relationships among numbers.
15. Can recognize benchmark numbers and number patterns.
16. Can recognize gross numerical errors.
17. Can understand and use equivalent forms and representations of numbers as well as equivalent expressions.
18. Can understand numbers as referents to measure things in the real world.
19. Can move seamlessly between the real world of quantities and the mathematical world of numbers and numerical expressions.
20. Can invent procedures for conducting numerical operations.
21. Can represent the same number in multiple ways depending on the context and purpose of the representation.
22. Can think or talk in a sensible way about the general properties of a numerical problem or expression—without doing any precise computation.
23. Engenders an expectation that numbers are useful and that mathematics has a certain regularity.
24. A non-algorithmic feel for numbers.
25. A well-organized conceptual network that enables a person to relate number and operation.
26. A conceptual structure that relies on many links among mathematical relationships, mathematical principles, and mathematical procedures.
27. A mental number line on which analog representations of numerical quantities can be manipulated.
28. A nonverbal, evolutionarily ancient, innate capacity to process approximate numerosities.
29. A skill or kind of knowledge about numbers rather than an intrinsic process.
30. A process that develops and matures with experience and knowledge.

Figure 1. List of components of number sense found in literature (Berch, 2005, p. 334)

Number sense goes much beyond the mere memorization and application of mathematical algorithms. As children are inclined to always draw on their natural insights and real-world experiences to derive a meaning for numbers and mathematical operations, they become convinced that mathematics makes sense and it is not a mere collection of rules to be remembered and applied (Howden, 1989). Because good number sense enables the child to view numbers as meaningful entities, they expect mathematical manipulations and outcomes to make sense (McIntosh et al., 1997). When the learner has such belief and expectations, they display a natural desire to look for links between new information and their prior knowledge and are inclined to prioritize the forming of connections between what they learn (McIntosh et al., 1997). Such inclination would naturally make learning more enjoyable and efficient and enhance a child's confidence and self-efficacy as a learner of mathematics and in general.

The paramount importance of number sense in children can be further seen in the close links it has with all the aspects of learning mathematics. Kilpatrick et al. (2002) have developed an outline of what successful mathematics learning is composed of, based on theories of cognitive psychology and mathematics education. They refer to this successful learning as "mathematical proficiency" and identify five interdependent strands of it (p. 9 – 16):

(1) Understanding- the learner comprehends mathematical concepts, operations, relations, as well as symbols, diagrams, and procedures.
(2) Computing- the learner is able to perform mathematical procedures such as adding, subtracting, multiplying, and dividing; does so in a flexible, accurate, and efficient manner.
(3) Applying- the learner can formulate problems mathematically and is able to devise strategies for solving these problems by choosing the mathematical concepts and procedures that are appropriate for the given context.
(4) Reasoning- the learner can rely on logic to justify the solution to a problem, or to expand from something known to something not known yet.
(5) Engaging- the learner perceives mathematics as sensible, useful, and doable.

A good number sense is at the foundation of all these mathematical proficiency pillars. As children develop insights on numbers and manipulations and how they are related to real world phenomena and representations, their *understanding* of mathematical concepts and operations deepens; drawing from real-life experiences makes them better able to conceptualize *computation* and gain fluency in it; the ability to form mathematical judgements and manage numerical situations makes pupils proficient in *applying* mathematical procedures to a given context; the inclination to look for links between prior knowledge and new information helps them enhance their *reasoning* during problem-solving; and the expectation that a procedure and result should make sense *engages* them to conclude that mathematics is sensible and useful. In fact, it has been long obvious that an underdeveloped number sense leads to “insuperable barriers learning mathematics” (Ekenstam, 1977, p. 317).

Dehaene (1997; see also Dehaene, 2011) has studied the origins of number sense in humans and other species, as well as how this ability is developed beyond its original capacity, drawing on the disciplines of neurobiology, neuropsychology, and cognitive psychology. Referring to the results of many tightly controlled experiments, Dehaene (1997, p. 13-14) concludes that many animal species are able to perceive numerical quantities without having been exposed to any training. This ability is not always perfectly accurate and it decreases with increasing quantities. The presence of such an ability can be explained by the forces driving the evolution of the species: perceiving numerical quantities helps animals compare two sources of food, or estimate the number of predators, skills that enhance their survival.

Dehaene (1997) reflects on the development of number sense research and the changes that happened in this domain around the 1980s. Before 1980s, when developmental psychology was dominated by Jean Piaget’s theory of constructivism, it was believed that the number concept in humans starts to be developed around the age of four or five. Many experiments developed by Piaget and his colleagues supported this claim, most notably the “number conservation” test (Piaget, 1952) where children are first shown two rows with the same number of coins spread equally apart and are asked “which row has more coins?”. Children correctly identify that both rows have the same amount. In the next step, the experimenter spreads the coins in one of the rows further, making the raw longer, while the other raw remains the same. When the experimenter repeats the same question, the child claims that the longer row has more coins, although the child witnessed that no coins were added or removed from any of the rows. It is not until the age of six or seven that children pass the number conservation test, realizing that changing the location or appearance of objects does not change their numerical quantity.

In the 1980s such constructivist claims were challenged by new experimental data and interpretations. It was claimed that like other species, humans also seem to be equipped with an inborn sensitivity to numerical quantities. Infant habituation experiments that exposed newborns or babies only a few months old to changes in quantity from two to three, showed that babies were able to notice and react to the change (cited in Dehaene, 1997, p. 49-50). Later experiments relying on neuroimaging of newborns have confirmed these results (Dehaene, 2011, p. 238). Other experiments with 3-4 year olds showed that while these children would fail the traditional Piagetian number conservation test, when these same children were shown two rows of M&M candies—one longer, but with fewer candies and one shorter, but with more candies—the children chose to take the row with the bigger amount, although they were not capable of counting yet. Interestingly, children of the age of two would succeed both in the M&M and traditional Piagetian test (cited in Dehaene, 1997, p. 45-47).

Such results prompted researchers to revisit Piaget's conclusions as assessors of a child's behavioral development, rather than his or her numerosity. For instance, Dehaene (1997, p. 46) explains the drop in performance in the number conservation test at the age of 3-4 with the development of "unconscious interference" in children of this age. Around this age children become able to go beyond the literal meaning of a sentence to retrieve the actual meaning the speaker is trying to convey. Due to this reason, when the experimenter repeats the same question "which row has more coins?" when the child has just answered a few seconds ago that "both rows have the same", the child thinks that the adult is not referring to the amount of coins anymore, but rather to something that has changed from the time when the child answered the question the first time. And since the only thing that the child has seen change is the length of one of the coin rows, the child concludes that the question is referring to that change and as such that must be the answer. Two-year olds on the other hand are not yet capable of such unconscious interferences, which might explain why they outperform their older peers at such a test. At the age of six or seven, the child becomes able to understand the testing situation better and realize that the question is the same, so they are not internally conflicted when giving the same answer twice.

But even researchers that see number sense as an innate ability in our and other species do not underestimate the important role that environment and experience have in developing this ability. In fact, there seems to be general consensus in research that such an ability can and should be fostered among children starting at a very young age. As Berch (2005, p. 336) has summarized,

> "Contrary to a strict nativist position, most theorists who adhere to the view that number sense has a long evolutionary history and a specialized cerebral substrate do *not* [emphasis added] judge that it thereby constitutes a fixed or immutable entity."

Instead, these elementary numerical abilities are seen as providing just the foundational structure for the acquisition of number sense, which needs to be developed in the environment through both formal and informal instruction (Berch, 2005, p. 336)

Many articles have been written on strategies that can be implemented within and outside of the classroom to enhance children's number sense (Gurganus, 2004; Aunio, 2006; Witzel et al., 2012). What is found in common in such research, is their emphasis for using concrete objects to help children explore numerical ideas and for relating mathematical procedures of more than, less than, grouping, decomposing, etc. with something that a child finds meaningful in his or her real-life experiences. Children enter pre-school with a good understanding of counting and approximation and have well-developed strategies to manage numerical situations, such as one-to-one matching, finger counting, etc. It is important that schools see these strategies as an asset, rather than a baggage they should quickly get rid of through formal instruction (Dehaene 1997, p. 139-140).

Experience shows that when initial teaching of mathematical algorithms is not coupled with the use of concrete and meaningful manipulatives that enhance a child's motivation and understanding, there is a danger of emerging innumeracy, the phenomenon of computing without thinking and as a result drawing false conclusions (Paulus, 1988). Stella Baruk (2016), a French mathematics educator has illustrated such a phenomenon by presenting first- and second-graders the problem "Twelve sheep and thirteen goats are on a boat. How old is the captain?" Many pupils answer "25 years old, because 12 + 13 = 25" prompted by the word "and" in the problem that they associate with the addition operation in mathematics. Dehaene (1997, p. 139-141) adds that in addition to the inadequacies of school instruction, the natural inclination of our brains to compartmentalize mathematical knowledge into multiple partially autonomous circuits is also an influencing factor leading to innumeracy. For him, the role of school mathematics is precisely to help the brain make meaningful connections between these individual brain compartments so that children can draw links between the mechanics of calculations and their meanings.

A good illustration of the compartmentalization of mathematical schemata is the early development of verbal counting and quantity comparison. Pre-school children might be fluent in verbally counting quantities, as they might be consistently correct picking out the group that has more items in it, when shown two different sets of objects. However, when asked the question "which one is bigger 5 or 7?" they would have difficulties to answer. Cognitive psychologists believe that such difficulties come as a result of verbal counting and quantity comparison developing in two separate parts of the brain and the learner does not connect the two yet, while in level 1 of number sense development. It in only in level 2, which is believed to be reached around the age of 6 years, triggered also by mathematical problems they face in school and home settings, that the child begins to elaborate on both of these schemata and map them onto each-other, becoming able to solve cross-modal questions involving the two. This leads them to develop the central numerical structure, that of the mental number line. In the next level 3, around the age of 7 children become capable to move across different number lines, those of counting by 2, by 5, by 10, by 100 etc. The final stage 4 of learning about the whole number is believed to typically start around the age of 9 or 10 years old, when both a general and an explicit understanding of the entire whole number system makes them capable in addition or subtraction with regrouping, estimation problems using large numbers, and mental math problems involving compensation. The development of children's conceptual understanding in mathematics is summarized in figure 2. (Kalchman, Moss, & Case, 2001, p. 3-4).

Level of Understanding	*Mathematical Domain*		
	Whole Numbers	*Rational Numbers*	*Functions*
Level 1: Consolidation of primitive schemata A: digital B: analog	A: Counting schema B: Qualitative quantity schema (more/less; addition & subtraction).	A: Formal halving and doubling schema, for numbers from 1 to 100 B: Qualitative proportionality schema, including visual halving and doubling	A: Recursive computation schema B: Bar graph schema
Level 2: Construction of new element A–B	Mental number line, with counting as an operation that is equivalent to addition and subtraction	Rational number line, with each number half or double previous one (e.g., whole, $^1/_2$, $^1/_4$, $^1/_8$, or we believe more appropriately; 100%, 50%, 25%, 12.5%)	Function schema, with line on Cartesian graph understood to represent results of iterative computation for different values of x
Level 3: Differentiation of new elements $A_1 - B_1$; $A_2 - B_2$	1s, 10s, 100s, and their relationship, understood and generalized to full whole number system	Decimals, fractions, percents, and their relationship understood	Confusable functions differentiated from each other, (e.g., $y = 2x$; $y = x^2$; $y = 2^x$. Function as object differentiated from function as sequence of operations
Level 4: Understanding of full system $A_1 - B_1 \times A_2 - B_2 \times A_3 - B_3$	1s, 10s, 100s, and their relationship, understood and generalized to full whole number system	Decimals, fractions, percents, and their relationship understood	Elements of polynomial (x, x^2, x^3) and the way they can relate, understood.

Figure 2 Development of conceptual understanding in three different mathematical domains (Kalchman, Moss & Case, 2001, p. 4)

Number sense is an often-used concept also in the research of mathematical learning disabilities (see for instance Gersten & Chard, 1999; Bachot et al., 2005; Wilson & Dehaene, 2007; Mazzocco et al., 2011). The purpose of such studies has usually been to establish an association between number sense and mathematical disabilities, such as dyscalculia (for example Mazzocco et al., 2011); the development of diagnostic tests that would measure different components of number sense which could predict dyscalculia and other mathematical disabilities (for example Baker et al., 2002); or the design of early intervention programs to develop number sense among young children as a way to prevent such learning difficulties from emerging (for example Aunio, 2006). Because this research concerns the development of number sense among pupils in the mainstream teaching, a deeper review of these studies is beyond the scope of this research. However, it is worth mentioning Dehaene's (1997) position that even though neurological pathologies may selectively impair mental calculation, they are infrequent. Although some researchers have estimated the percentage of "mathematically disabled" children to be around 6%, in Dehaene's opinion this is a largely overstated amount and some of these children are being misdiagnosed. What he believes to be the mechanism, is that some pupils might have gotten the wrong start in mathematics, convincing them that this is a domain they will never be able to make any sense of. This emotional component of anxiety or phobia about mathematics makes them consistently underperform in the subject (Dehaene, 1997, p. 140-141).

To conclude, the numerosity and diversity of definitions for number sense shows the wide attention it has received in mathematics education research, as well as how complex and multifaceted the ability is. The definitions are however not in direct contradiction with one another—instead they simply capture the many diverse ways of how number sense manifests itself in the mathematics learning. There seems to be general agreement among cognitive psychologists and educational researchers on both the innate origins of number intuition in the human species, and the need for high quality formal and informal instruction to develop this intuition to the level needed for further conceptual understanding and learning. Development models of number sense as a function of age have also been developed, suggesting the full understanding of the whole number domain to happen typically by the age of 10, whereas for rational numbers and functions at later ages.

2.2 Judging the reasonableness of computational results

The ability to judge the reasonableness of mathematical statements and results is related to whether a child can assess if a result is correct or not, based on its meaning in the real world or whether the result meets or contradicts the expectation that the child had formed about the result. For instance, a child with a good understanding of magnitude for big numbers would conclude that the statement "5000 textbooks can fit into my schoolbag" is not correct, as it contradicts what is possible in the real world. Similarly, if a child performed a mathematical algorithm and found out that "149 x 4 = 5636", he or she might be surprised by the large result, if they expected the answer to be less than 800, since 149 is less than 200 and 200 x 4 = 800. The child would then revisit the computation and probably notice the mistake in the place-value that occurred during the original computation.

Case (1998, p.1) described number sense as something that is "difficult to define, but easy to recognize." In research number sense has been continually operationalized and measured using its identified comprising components. Kalchman, Moss, & Case (2001, p.2) have reviewed how a number of authors have operationalized number sense and based on that have identified that:

> "the characteristics of good number sense include: (a) fluency in estimating and judging magnitude, (b) ability to recognize unreasonable results, (c) flexibility when mentally computing, (d) ability to move among different representations and to use the most appropriate representation for a given situation, and (e) ability to represent the same number or function in multiple ways, depending on the context and purpose of this representation."

The ability to judge the reasonableness of computational results engages at least four out of these five abilities listed above. When judging which of the provided results is the most reasonable to answer a certain question or computational result, the pupils need to estimate the magnitude of the expressed quantities (i.e. 5 dl vs. 5 l, etc.); recognize unreasonable statements (i.e. 5000 textbooks cannot fit into a schoolbag); perform mental computations (i.e. 149 x 4 is less than 600 because 150 x 4 = 600); and move between different representations (i.e. 300 cm is 3 m, etc.). This might explain why many researchers assess children's number sense by measuring their performance in judging the reasonableness of a computational result (Yang, 2019; Alajami & Reys, 2010; NCTM, 2000; McIntosh et al., 1997).

Indeed, number sense enables pupils to view mathematics as meaningful and to expect operations and outcomes to make sense. So they continually rely on a variety of internal “checks and balances” to judge the reasonableness of numerical outcomes. When such an outcome conflicts with the expectation, the child revisits the mathematical situation and attempts to solve the conflict, usually by correcting the operational algorithm or choosing a more appropriate solution strategy altogether. So number sense exhibits itself in a number of ways as the learner engages in mathematical thinking, including being aware of accuracy and sensitive to the reasonableness of calculations. (McIntosh et al., 1997).

Naturally, the ability to judge the reasonableness of a computational result leads to higher accuracy in performing mathematical operations. As pupils are sensitive to unreasonable and incorrect results, they will likely re-perform the algorithm or change the solution strategy to derive the accurate answer. This enhances their performance in mathematical problem-solving both in and outside the classroom. Additionally, this ability makes pupils better learners, as they continually look for other solution strategies and draw meaningful links between these different strategies and different operations. For instance, they might gain deeper understanding how addition and subtraction are related, or how fractions and decimals are different representations of the same quantities, etc. Besides making them better learners and performers, the ability to judge the reasonableness of computational results influences also how children perceive themselves as learners of mathematics. According to Howden (1989, p.7):

> “Students who can make judgments about the reasonableness of computational results and realize that more than one way can be used to arrive at a solution gain confidence in their ability to do mathematics. Research has shown that such confidence influences students’ view of themselves as mathematics learners and their future decisions about studying mathematics.”

Given the multifaceted importance such an ability has in mathematics learning, it is not a surprise that the mathematics curriculum also emphasizes judging the reasonableness of computational results. The National Council of Teachers of Mathematics (NCTM, 2000) has emphasized that school mathematics should make learners proficient in using estimation as a tool to perform computations already in early-childhood and primary education, and further make pupils in grades 3-12 fluent in employing a variety of strategies to judge the reasonableness of numerical computations and their results. The NCTM principles and standards for school mathematics inform to different extents the national mathematics curriculums worldwide.

The Finnish National Core Curriculum for Basic Education (2014) has also emphasized the development of the skill to judge the reasonableness of computational results in school mathematics. Among the fourteen central objectives of instruction in mathematics in grades 3-6, it is listed "O6 to guide the pupil to develop his or her skills in assessing whether the solution is reasonable and meaningful" (p. 252). Under the key content area C2 of numbers and operations, the curriculum states that the pupils should be "guided to round up figures and to calculate with approximate values, through which they learn to estimate the order of magnitude of the result" (p. 253), whereas under area C4 of geometry and measuring, pupils should "practice measuring and pay attention to the accuracy of measurement, estimation of the measurement results, and verifying measurements" (p. 254).

Such an emphasis of the ability to assess the reasonableness of the solution is new to the Finnish curriculum. The Finnish National Core Curriculum for Basic Education 2004 does not include the development of such an ability as a learning objective and the only mentioning of something related to it can be found in a list of core contents for "Numbers and Calculations", where one of the thirteen list items is "evaluating, checking and rounding the results of calculations" (p. 161). The checking and rounding of the calculation results are generally understood as to be performed through mathematical operations and the term "evaluating" in this sentence lacks both clarity and emphasis. The newer curriculum of 2014 does a much better work in clearly emphasizing the ability to assess the reasonableness and meaningfulness of derived solutions to mathematical problems. Other national curriculums worldwide are lagging behind. Yang (2019) has problematized the fact that Taiwan is one of the many countries that has failed to include judging reasonableness of solutions in the national curriculum.

The fact that this ability is missing from the curriculae partially explains why judging reasonableness is an area that gets the least attention in the teaching of mathematics. In many countries the skills is missing from both pre-service and in-service teacher training, leading to teachers not being of what the concept means nor on how to develop or assess it. Another explanation stems from the nature of mathematics as an exact science, where "reasonable" is not good enough of a solution—it is the only one right answer that we should seek for.

Interviews with middle-school mathematics teachers in Kuwait revealed both of these reasons (Alajmi & Reys, 2007). Twelve out of the thirteen interviewed teachers interpreted "reasonable answers" to mean something very different from how it is defined in the mathematics education literature. Half of these teachers thought that reasonable answers were the exact, correct answers, since mathematics is an exact science and you cannot possibly have two answers to a mathematical problem—an answer is therefore either correct and reasonable, or wrong and unreasonable. The other half thought reasonable answers were those that were derived through mathematical procedures that were (at least partially) correct, even if such answers deviated from the correct answer. For instance, one teacher gave as an example that if a pupil concluded that “3/5 + 1/3 = 4/15”, this could be considered a reasonable answer, because even though it is wrong, the pupil has correctly identified the common denominator of 15 and that the result would be a fraction of this denominator. However, considering the meaning of reasonable answers in the mathematics education literature, a child with good number sense would identify this as an unreasonable answer, reasoning that “5/15 is 1/3, so 4/15 is less than 1/3, so it must be the wrong answer, because if I add a positive number to 1/3, I should get a number that is bigger than 1/3 and not a smaller one”. Only one of the thirteen interviewed teachers displayed a similar understanding of “reasonable answers” as the one found in literature. Mostly teachers did not view this as an important teaching area, as it was not mentioned in the curriculum.

To sum up, the ability to judge the reasonableness of computational results is considered an important component of number sense and has been often used in research as a proxy for assessing number sense. Education researchers view the ability to judge the reasonableness of obtained answers as a very important skill in mathematics learning, as it prompts the learner to refer to the priorly accumulated knowledge and understanding during the solving of a new problem, leading to more effective learning; it enhances the chances of noticing mistakes and, when needed, of revisiting solution strategies, improving performance; and it increases learner confidence in mathematics. Despite this long established importance in the research domain, many curriculum setting circles have failed to include the judging reasonableness of computational results in the mathematics learning objectives and (consequently) teachers seem to lack awareness on what it is, or how to teach it. The Finnish National Core Curriculum for Basic Education 2014 has listed the ability to assess whether a solution is reasonable and meaningful among the central objectives of mathematics instruction.

2.3 Misconceptions

Evolution has equipped humans with the capacity to develop explanatory mental models, or constructions in an attempt to make sense of the world around them. In devising such constructions people reflect and conclude based on their past life experiences. When introduced to a new fact or concept, the learner needs to fit this new piece of information into a previously constructed model, or otherwise the new concept will not be recalled for future use. These cognitive processes that are employed in our learning of science are captured by constructivism. The constructivist model predicts that once a new concept has been explained to a class of thirty pupils, at the end of the lesson each one of these pupils will have constructed their own version of that concept that differs, at least slightly, from everybody else's, since each learner has an individually unique set of past experiences and pre-constructed models to relate this concept to. When these existing constructions are at odds with accepted science they lead to incorrectly constructed scientific concepts, known as misconceptions. (Allen, 2014, p. 24-30).

Misconceptions are therefore a sign of both our natural inclination to learn and at the same time our limitations in doing so. In the quest to explain the world around us before we have all the facts, we reach incorrect conclusions. However, as emphasized in David Ausubel's (1968) theory of meaningful learning, these pre-maturely built structures are important for learners to absorb new information. When children link the new knowledge to ideas they already had in place, they develop more active and more meaningful learning strategies and are capable of independently looking for answers by searching for general principles that connect their isolated pieces of scientific knowledge. For Novak (2002), meaningful learning takes place in a continuum depending on (i) the amount, quality and organization of the prior knowledge, and (ii) on the efforts of the learner to integrate the new information into the existing knowledge framework (figure 3).

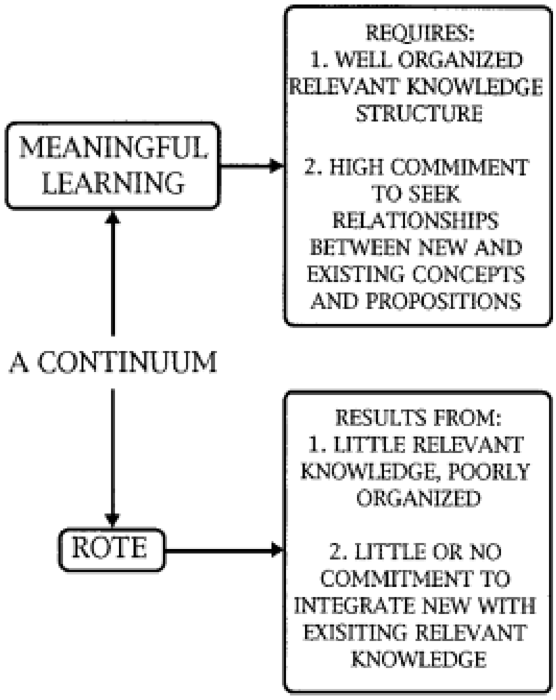

Figure 3 The influencing factors in the continuum of meaningful learning (Novak, 2002, p. 552)

Natural and important as they are, misconceptual constructs interfere with successful learning, when left unidentified and not replaced with correct scientific information. Piaget (1955) studied and wrote extensively on how children progress through the cognitive developmental stages by correcting constructs they had built in the prior stage. For him, children should be asked questions to make their constructs explicit and then be introduced to experiences and/or that contradict these constructs. The child's mind will result in an internal conflict between the existing construct they have and the new experience they are exposed to. The attempt to solve this conflict leads them to replace their misconceptual construct with more correct knowledge that can accommodate the new information. Vygotsky emphasized strongly the role of both adults and peers in providing the learner with new experiences and helping them overcome their intellectual conflict in the zone of their proximal development (cited in Daniels, 1996).

Drawing on these theories it becomes clear that the teacher plays an important role both in identifying misconceptions and correcting them to give space to meaningful learning. Pupils often bring into the classroom certain misconceptions formed over early development years when they were not yet introduced to any science concepts. For instance, common misconceptions on how forces behave are generally formed during informal play in early childhood. Other times, pupils may form misconceptions during the lesson itself, as they interpret the new information they get in the light of their past experiences and existing constructs. An educator could become aware of both these kinds by reading literature that summarizes the so-called "classic" misconceptions children have in primary science, and more importantly by actively engaging pupils in the elicitation of their conceptual frameworks, a process during which learners become explicitly aware of what they already believe about scientific phenomena. (Allen, 2014).

Allen (2014) lists a number of strategies that teachers could use to elicit pupils' conceptions, be they right or wrong, such as: asking pupils to make different explanations and predictions, giving them self-completion exercise worksheets, engaging them in card-sorting activities, paying attention to pupil's drawings, asking pupils to complete and explain concept maps (usually after a topic has been delivered), using concept cartoons (different cartoons offer different explanations and the pupil has to decide which one is correct), observing children's play and role play, engaging in word-association games, and eavesdropping on their conversations during group work, etc.

Strategies on correcting misconceptions however are much less straight-forward and success is hardly guaranteed. Novak (2002) states that hundreds of studies on misconceptions prove that "facilitating student's acquisition of powerful and valid conceptual frameworks is not easy" and warns that "there are innumerable ways to go wrong and no set of traditional instructional strategies that are foolproof" (p. 555). It seems like the resources we have go back to Piaget's and Vygotsky's thinking of exposing the child to meaningful new experiences and engaging them in conversations that help them adjust their existing constructs to advance to a new stage of cognitive development.

Ryan & Williams (2007, p. 13 - 30) classified the errors made by children aged 4-15 in mathematics standardized assessment and national tests in Great Britain and after excluding the mistakes resulting from "slips" and those of an "uncertain diagnosis," they identified four main types of misconceptions that hinder children's learning in mathematics. These misconceptions were related to modelling, prototyping, overgeneralizing, and process-object linking. Below there is a short explanation for each.

Modelling refers to the way mathematics is connected to the real world. A modelling error as such refers to a situation where the child has created his or her model of the mathematical problem or task, and this model differs from the mathematical model that was expected in the academic context; or the child has challenges in finding a real-life model that would correspond to the given task, which then prevents him or her from completing the task. For instance, a preschooler is usually able to answer to the question "I have two bricks here and I take one more brick. How many bricks do I have now?" even when there are no actual bricks being used in the situation. However, the same child might have difficulties answering to "what is 2 add 1?" Models are important to bring meaning to mathematics by making connections between what is intuitively known and mathematics concepts and operations. However, they provide also limitations, as usually there is a gap between a child's use of the model and target mathematics that pedagogy is yet to address (Ryan & Williams, 2007, p. 19).

Prototypes refer to "culturally typical example of the concept" (Ryan & Williams, 2007, p. 20). For the human brain it is much more natural to learn a new concept not in the mathematical way, but in the prototypical way. This may lead to reading and perceiving data incorrectly, or drawing the wrong conclusions. For instance, when hearing the concept "rectangle" the child has often in mind a prototypical example of a figure with four right angles where two of the sides are longer than the other two. Such an image may lead them to not recognize a square as a rectangle, although in the mathematical world squares are a subset of the rectangle set. Similarly, because divisions in between two quantities usually indicate an increase by one—like in the case of a rules, where each subsequent line means one millimeter more—a child may read incorrectly a graph that has used a less-common scaling of incrementing by two or four units at a time. Misconceptions spanning from prototyping should be addressed by exposing children frequently to situations that challenge these typical prototypes they may have.

Overgeneralizations explain most of the misconceptions in mathematics held by children between the ages 8-16 years old. Overgeneralizations occur when children apply rules that they have found valid in one domain to other domains where such rules do not apply anymore. The most typical overgeneralizations are those of expanding rules of operations with whole numbers into the domains or fractions, decimals and negative numbers. For instance, children might expect the operation of multiplication to always yield a bigger result than the multiplicands itself, because the operation of multiplication makes (usually) whole numbers bigger. This is however not true in the case of fractions, decimals, or negative numbers. Or pupils tend to conclude that 0.5 is smaller than 0.32 since 5 is smaller than 32. Teachers can help the correction of such misconceptions by purposefully asking pupils to explicitly state some generalizations that they are inclined to make and asking them to bring up examples when this would not be true.

Process-object linking is related to the conceptual understanding of what mathematical processes, often captured by operations and solution strategies, are related to what mathematical objects, often related to the answer to a problem. For instance, in early development when children are asked “how many buttons are there?” the child will count the whole set of buttons. The object that is sought for (in this case the answer to how many buttons, or in mathematical terms the cardinality of the buttons’ set) is related in the child’s mind to the process of counting, rather than to the last number they stated out loud when pointing at the last button. So if you ask the child that has just finished counting “so how many are there?” the child starts counting again, due to this incorrect link between the process and object. Conceptual changes asking pupils to derive this “reification” between process and object occur throughout the learning of mathematics. Teachers can help pupils build meaningful links between processes and objects in mathematics by re-introducing models and making connections between different models.

In general, Ryan & Williams (2007, p. 27-28) urge educators to investigate children’s errors in mathematics, ponder upon the types of misconceptions that have led to such errors, and design effective pedagogies to correct these misconceptions. A mere correction of the error, they warn, would be harmful as it may lead the child to conclude that mathematics makes no sense and is an arcane activity.

2.4 Findings of prior research

Although number sense has gotten abundant attention both in the research community and in the mathematics curriculum setters’ circles, studies focusing specifically on the assessment of number sense across school grades has been missing in the European and American setting. Instead, Western studies on number sense have focused mostly on the development of number sense, its predictive power in relation to future mathematical proficiency or learning disabilities in mathematics, intervention programs to develop number sense, etc. Some influential studies in the field are mentioned in section 1.1.

This section focuses on studies that deal specifically with the measurement of number sense in basic education grades 1-9. The only exception are Pirjo Aunio’s (2004; 2006) studies, which focus on the assessment of number-sense among the pre-school age group. The reason behind this exception is the intention to include the Finnish research landscape in the domain of number sense, while no studies that assess number sense in school age groups have been conducted in Finland.

Pirjo Aunio has examined the influence of nationality, age, and gender on the development of number sense among pre-school children in Finland, Hong Kong, and Singapore (Aunio et al., 2004) as well as in Finland and China (Aunio et al., 2006). Researchers analyzed the performance in an early numeracy test of 130 Chinese children and 230 Finnish children aged between 4-7 years old. Both samples showed a clear increase in number sense as a function of age, both in counting and relational skills (i.e. the ability to organize and compare quantities). A comparison between the two tests showed that Chinese children outperform Finnish children in counting skills at all age categories. In relational skills, Chinese and Finnish children performed the same in the lower ages, whereas a difference was noted among the older children, where Chinese performance was slightly higher. In addition to cultural differences in the academic ethos and setting, researchers explained this difference in performance also with the language—names for numbers in Chinese and other Asian languages are much more meaningful than in Western languages. For example eleven in Chinese, is ten-one making it easy for the child to operate in the base-ten system. The Finnish equivalents for numbers from 10-20 are not as logical. In the comparison of Finnish, Singaporean, and Hong Kong children Aunio et al. (2004) found that children in Singapore performed the best and Finnish children had the lowest performance among this group in both counting and relational skills at all age levels. Differences in teaching were assumed to explain these results. No gender differences were found in any of these studies.

The pioneering study in measuring number sense among school children of different ages was conducted in 1997 in four different countries: the United States, Australia, Taiwan, and Sweden (McIntosh et al., 1997). After that, assessing number sense seems to have been much more pertinent in the Asian setting, where different components of number sense have been measured in Taiwan (Reys & Yang, 1998; Yang, Li, and Lin, 2008; Lin, Yang, & Li, 2015; Yang, 2019), Kuwait (Alajmi & Reys, 2010), and Malaysia (Mohamed & Johnny, 2010). A more detailed review of these studies in the chronological order follows, concluded with an overall summary of their general findings as well as a critical stand on this research at the end of the section.

McIntosh et al. (1997) study assesses the general level of number sense in students of different ages (8, 10, 12, 14) in four different countries, Australia, Sweden, Taiwan, and the United States to examine whether number sense does develop with more schooling experience. The purpose was not to compare the performance in these four countries, as in Taiwan and Sweden not all the ages were tested and not all the questions of the test were used. Researchers intended rather to gain a better understanding of the development of number sense by measuring performance in different context and academic settings. The findings showed that number sense did indeed develop with the age of the participants. Further researchers found that developing technical expertise in pen-and-paper computations does not necessarily contribute to the development of number sense.

Prompted by consistent research results that showed Asian children superior in mathematical skills compared to their Western peers, Reys & Yang (1998) aimed to examine whether this superior performance was confined to computational skills, or it extended itself to possessing a better developed number sense. They concluded that while Taiwanese sixth and eighth graders performed very high in written computation, their performance was significantly lower in similar questions that relied on their number sense. Researchers found little evidence that students would rely on number sense components in problem solving. They called for more caution when reading comparative test results that measure only the percentage of correct answers.

Bana & Dolma (2006) compared students' computation and estimation skills, where estimation was considered as a proxy that measured number sense. They tested 77 seventh graders in Western Australia using fifteen matched questions where one tested computation skills and one estimation abilities in all the three areas of whole numbers, decimals, and fractions. Results showed students' estimation aptitude to score at 41%, ten percentage points lower than their computation skills. The most problematic areas were those of fractions and decimals, where students held significant misconceptions.

In 2008, Yang, Li, and Lin studied (i) whether there were significant differences among different components of number sense, meaning if students performance across different areas that measured number sense varied significantly; (ii) whether gender had an effect on such a performance; and (iii) if the development of number sense and achievement in mathematics were significantly correlated. After testing 1,212 Taiwanese pupils who had just completed the fifth grade curriculum, they found that pupils perform best at "recognizing relative number size" and worst at "judging the reasonableness of computational results"; female students perform slightly better than their male peers in recognizing relative number size; and mathematics achievement of students is highly correlated with how well their number sense is developed.

Mohamed & Johnny (2010) investigated the relationship between number sense and mathematical achievement and explored the components of number sense that are most problematic for students. After testing 32 Malaysian fourth graders that had achieved a score of 80% or higher in the national annual mathematics test, researchers reported that despite the high achievement with a mean score of 86,38% in the test, the same sample of pupils only recorded a mean score or 58,28% in number sense competency. The most problematic areas were those of recognizing the relative effect of operations on numbers and judging the reasonableness of computational results.

Alajmi & Reys (2010) examined the performance of 200 eighth grade students in Kuwait in recognizing reasonable answers, as well as the strategies they used to determine reasonableness. They found that 35% of the students relied on their number sense to identify reasonable answers, using as a criteria either (i) the relationships of numbers and the effect of operations, or (ii) the practicality of the answers. The strategies they used included estimation, numerical benchmarks, real-world benchmarks, and their understanding of properties of mathematical operations. Over 60% of the sampled students relied on algorithmic techniques and pen-and-paper calculations to identify reasonable answers. Generally, students performed low in judging the reasonableness of answers across all three number domains of whole numbers, fractions, and decimals. The authors advocated for a shift in mathematical teaching that would pay more attention to computational estimation and the development of number sense among students, instead of having an abundance of performing pen-and-paper algorithms.

Lin, Yang & Li (2015) developed and used a web-based two-tier test to assess number sense in a sample of 1,248 sixth graders in Taiwan. Pupils were tested in five number-sense component areas, as defined by the researchers: (i) understanding the meaning of numbers across whole numbers, fractions, and decimals and the use of multiple ways to present numbers; (ii) recognizing relative number size, for instance when comparing fractions; (iii) ability to switch among different representations of quantities and choosing the most appropriate representation for the context, (iv) recognizing the relative effect that operations have on numbers, i.e. does an operation make the number larger or smaller, and (v) being able to judge the reasonableness of a computational result. The average percentage of correct answers obtained across all the five areas was about 45%, among which about 22,9% relied on number-sense methods to solve the problems. From all the identified five areas of number sense, students performed the lowest in judging the reasonableness of a computational result, where the percentage of correct answers was 36%, among which 17,5% relied on number-sense strategies.

Yang (2019) study focused on the ability to judge the reasonableness of a computational result and assessed this ability in 790 fourth graders using a web-based two-tier test. Findings for the first-tier (answer-tier) questions showed a correct answer percentage rate between 31,8 – 60,8%

for all the eight questions asked, with the total average for all the items being 51,7%. Answers to the second-tier (reason-tier) questions revealed that only 25% of the sampled fourth-graders relied on number-sense methods to answer the first-tier questions, whereas about 38% of the pupils had misconceptions pertaining to number magnitude, or the effect of operations on numbers.

A cross-sectional view at all these findings leads to a few important conclusions that have been supported continuously by research: (i) school children's number sense is less developed than their computational skills; (ii) this result holds true also among pupils/countries that are top-performers in standardized mathematics tests; (iii) the emphasis of mechanical computations over number sense in the national mathematics curricula, textbooks and teacher attitudes may explain why number sense lags significantly behind computational skills; (iii) the component of number sense where pupils perform the lowest is that of judging the reasonableness of computational results; (iv) no significant gender differences are noticed in number sense development.

While the drawing of such general conclusions is easy due to the lack of controversy in the findings, such results must be interpreted with a grain of salt. First, the amount of such research is quite limited. Second, the geographical representation of the research settings is limited, as certain regions or countries are over-represented and most of the academic settings globally are missing all together. Moreover, it is clear that all the researchers behind these studies have extensively collaborated with and influenced each-other's work, which might potentially lead to a certain degree of academic bias on the phenomenon. The content of the instrument tests is, for instance quite similar, and the tone with which the results are reported sways towards a slight dramatization of the problem. It is clear that more studies from other academic settings and new researchers using a wider range of diversified assessment instruments are needed to draw more reliable conclusions on the extent to which number sense is developed among different stages of primary education.

3 Data and methodology

This section gives an overview of the collection and analysis of the data used in this study, providing descriptions and relevant justifications for the test instrument that was used, the sample of pupils, and the way the data was collected and analyzed. Related methodology limitations, as well as the attempts to address these methodological pitfalls are also discussed when relevant.

Instrument

Researchers have considered the assessment of number sense as a challenging task for researchers and educators alike (Durkin & Rittle-Johnson, 2015). Until 2015 the measuring of pupil's number sense was conducted through pen-and-paper tests that recorded the percentage of right versus wrong answers, and follow-up interviews with a small sample of selected pupils to identify the reasoning that took place during problem-solving (for example McIntosh et al., 1997; Reys & Yang, 1998; Bana & Dolma, 2006; Alajmi & Reys, 2010, etc.). This data collection method took a lot of time from both pupils to answer and researchers to analyze, making it a not-so-effective way for larger scale studies. Furthermore, as interviews were conducted only with a small number of participating pupils, valuable information on how the rest of the pupils reasoned was not captured by the studies.

To address these challenges that pen-and-paper tests followed by selected pupils' interviews have in assessing number sense and identifying related misconceptions, Lin, Yang & Li (2015) developed a web-based two tier test. The first tier (answer-tier) assesses the content knowledge of the respondents and gives information on the relative weight of right versus wrong answers. The second tier (reason-tier) aims to identify the reason for the first-tier response by capturing the respondent's thinking process. While two-tier tests had been used in science education, in mathematics they were less emphasized until this point, probably due to the challenge in designing the reason-tier, since there exist multiple approaches in solving mathematical problems (Lin, Yang, & Li, 2015). The authors recommended relying on literature that identifies a certain age group's thinking process and misconceptions in a given mathematical domain, when designing the reason-tier of the test. After its publication, the test has been considered both efficient and convenient in assessing pupils' number sense performance, identifying their reasoning during the solving of problems, and showing some of the misconceptions pupils have (Yang, 2015). Similar two-tier tests have since been used to assess different components of number sense among pupils (Yang & Lin, 2015; Yang, 2019).

For this research, a similar web-based two-tier test was used as an instrument to evaluate pupils' performance in one component of number sense, that of judging the reasonableness of computational results. The test was adapted from the instrument used by Yang (2019), who formulated the test based on a review of the number-sense literature, as well as previous two-tiered tests used in science education research. Basing the test on a successfully-tried out instrument in good-quality international research enhances the credibility of the research, as well as provides a ground for comparing these research results to prior findings. Furthermore, as proven in prior studies, this test is an effective tool both in assessing number sense components and in revealing frequent misconceptions pupils might have in this domain (Yang, 2015).

The Yang (2019) test content was modified in line with the learning objectives from the Finnish National Core Curriculum for Basic Education (2014), as well as the local curriculum and learning materials used in the school that participated in the research. As a result of these modifications, measurement units of milliliters (ml) and millimeters (mm) were removed from the questions' multiple choice alternatives. Additionally two questions related to operations with fractions (question 9) and decimal numbers (question 10) respectively were added to the original test questions.

The translation of the test into Finnish language was checked by a professional translator—for language structure and grammar, by a mathematics student teacher—for field terminology and concepts, and by two student teachers that had worked with the tested group of pupils—to ensure the language and mathematical symbols used would be familiar and suitable for the level of the test audience. The test was further shown to the four classroom teachers of the groups that participated in the test and was considered by these teachers to be a suitable instrument for its designated purpose. One pupil of the same groupage and grade level as the tested children took the test prior to the data collection, to confirm the format, concepts and notations of the questions were understood, as well as to measure the time needed to answer.

The test was designed in two tiers. The first-tier multiple-choice questions 1-10 assessed the pupils' number-sense performance by measuring their ability to judge how reasonable a statement or computational result was. Based on the alternative that the pupil had selected to answer each of these questions, a second-tier question—also multiple-choice—appeared to ask the pupil to explain their reasoning when answering the question. For example, in question three pupils were asked:

3. Emil is 10 years old and 130 cm tall. How tall will Emil most likely be when he is 20 years old?
    a. 170 cm
    b. 260 cm
    c. 350 cm
    d. I cannot tell

If the pupil's answer was alternative A, the second-tier question that followed was:

3A. I answered this way because:

- i. It is possible for a person to be 170 cm tall.
- ii. 260 cm and 350 cm are too tall, so the correct answer must be A.
- iii. I guessed.
- iv. Other______________________________

After answering this question, the computer would automatically direct the pupil to first-tier question number four. Whereas for a pupil that chose alternative B in question three, the second-tier question would be:

3B. I answered this way because:

- i. Because Emil will be two times older (2 x 10 = 20) he will also be 2 times taller (2 x 130 cm = 260 cm)
- ii. 170 cm is too short and 350 cm is too tall, so the correct answer must be B.
- iii. I guessed.
- iv. Other___________________

The second-tier questions aim to explain the pupils' reasoning in solving problems pertaining to number sense and potentially to make explicit any misconceptions the pupils might have in the area. The list of the complete English questions in both tiers can be found in the respective tables in the results section, whereas the full test in Finnish is attached in Appendix 1 of this study.

Sample

A total of 90 fourth-grade pupils (aged 10-11 years old) from 4 different classes of a school in Northern Finland participated in the research. This age group was chosen having in mind the current development model of conceptual understanding in mathematics, according to which by the age of 10 children typically have formed a full general understanding of the whole numbers system, enabling them to perform mental arithmetics and estimation problems (see Kalchman, Moss & Case, 2001 in the theoretical background section). Furthermore, the test developed by Yang (2019) was designed for the same age group. The school was chosen due to its relatively large size and its affinity for education research. The school is divided into two different campuses located in different parts of the city, providing for a diverse socio-economic background of its pupils. Furthermore, the pupils of the school have been priorly exposed to participation in education research and are accustomed to taking tests in electronic devices such as computers and tablets.

Data collection

The permission for collecting the data was approved by the principal of the school and the teachers of the participating classes. Pupils' parents were also sent the research permission details and were given the option to have their child opt out of participation. Pupils were also explained that participation in the study was voluntary and their answers were fully anonymous and as such would not affect their evaluation in the subject. All the pupils that were present the day of the test participated in the study. Pupils were explained the purpose and structure of the test by their teacher and the researcher. During these instructions, the word "test" was avoided not to cause any stress feelings related to exam-taking. All pupils seemed very relaxed and at ease during and after they had answered.

Pupils answered the test in their individual iPad-devices, which are very familiar to them without using pen, paper, calculator or any other aiding devices. Pupils were allowed to take the test for a whole class period of 45 minutes, but all of them had finished before this time. The average time to complete the test from all the respondents was about 15 minutes. At the end of the test, all pupils received in their individual devices an encouraging message on their performance and were not given points based on their performance. Instead, they were told that the questions and answers would be sent to their teachers and could be discussed at some later occasion, if there was interest and time to do so. Pupils' responses were collected using a Microsoft form in the O365 cloud.

Data analysis

For the 10 first-tier questions, the responses were classified into correct and incorrect and the corresponding percentages for both categories were calculated and reported (table 11). The answers to the second-tier questions were manually categorized based on the method that the pupil had used in deriving the answer. The main categories were those of a number-sense method (N), a rule-based method (R), misconceptions (M), and guessing (G). In addition to these four main categories, additional categories of "more information" (MI), "pen-and-paper calculations" (PC), and "challenges in understanding the terminology" (T) were created only for questions 1, 9, and 10 where the frequency of such items was high enough (over 5 answers per category). These categories corresponded to pupils' answers such as "I need more information to answer" (MI), "I need to make calculations with pen and paper to answer" (PC), and "I do not remember what [a concept used in the question] means" (T). Having these additional categories gives valuable information on pupils' reasoning and it avoids a situation where their answers would be misclassified into one of the four main categories. In statistics it is recommended that when the Chi-squared distribution modes is used, categories are constructed such that the frequencies are not too small—a common cutting point is that of at least 5 observations in each category (Lindsay, 1995, p.87). In question 2 the frequency of answers in each of the MI, PC, and T categories did not make the cutting point of 5, so for analysis purposes these three categories were grouped together under "Other" (O). Similarly, categories of rule-based method (R) and number-sense method (N) were removed from the analysis in questions 3 and 10 respectively due to either a lack of such category in the provided answer alternatives (question 3) or a too low frequency of less than 5 answers in this category (question 10).

After the categorizing, the frequency of second-tier answers in each category was counted and reported as a percentage (table 11). Subsequently, a chi-squared test for categorical data was performed using the Excel software program for each of the 10 first-tier questions, to test whether the distribution of the N, R, M, G and other categories in each question differed in a statistically significant way.

The Pearson chi-squared test is used to determine whether the observed frequencies differ significantly from the expected frequencies across several pre-defined mutually-exclusive categories (Plackett, 1983). In this research, this test shows whether pupils rely significantly more on misconceptions than on number-sense or rule-based methods when answering a certain question, for example. One limitation of the test, is that it is not reliable if the frequency of observations in a certain category is less than 5 observations (Lindsay, 1995, p.87). For this reason, the analyzed categories in each question have a frequency between 5-69 observations. Corrections that could be made to the test to make it applicable to categories with less than 5 observations, such as the Yate's continuity correction were avoided, as they hinder the reliability of the test (Hitchcock, 2009). Another limitation is that the test does not provide a tool for assessing how sensibly the mutually exclusive categories have been identified, or whether observations have been correctly placed into these corresponding categories (Lancaster & Seneca, 2005). To address this limitation, the categories used in this research are aligned with those found in prior published studies on the same topic (i.e. Yang, 2019).

4 Research results

Results of the data analysis are summarized in this section. The percentage distribution of answers for both first- and second-tier questions are shown in respective tables and a discussion follows for each table. Respective results from the Yang (2019) study are also mentioned for the relevant questions, although the aim is not to make a direct comparison of such findings, since the pupil samples and the data collection setting in these two studies differed to an extent that would make such a comparison not meaningful.

The section concludes with a general summary of the percentages of the correct answers for each first-tier question and the respective frequencies of the problem-solving strategies identified in the second-tier questions (table 11). The chi-square test results showing whether statistically significant differences are noted among the frequencies of the different strategies are also included in table 11. An even broader discussion of the findings and of how they relate to prior research follows in the conclusion section of the thesis.

4.1 Pupils' responses to test questions 1-10

Pupils' responses to all the test questions 1 – 10 are analyzed below. I begin by explaining what mathematical thinking skills the question involves and then I analyze the obtained answers, focusing especially on the number-sense problem solving method and the related misconceptions pupils seem to have for each question item.

4.1.1 Pupils' responses to test question 1

Table 1 summarizes the first- and second-tier pupil answers to question 1. The aim of this question is to assess whether the pupil understands how the same amount of 5 000 is or is not reasonable, when pertaining to different daily life situations. So, while 5 000 textbooks is an unreasonable amount to carry, 5 000 g on the other hand is sensible. The answers reflect a child's ability to grasp quantities represented by big numbers and to understand how the context may change the meaning of the same quantity.

Table 1 Sampled pupils' responses to question 1 on the first- and second-tier tests

1. Whose statement is the most reasonable:			
Emmi: I can fit 5 000 textbooks in my schoolbag.	1,1 %		5 000 textbooks is not that many and a school bag can fit them.
		1,1 %	5 000 g is too heavy and the mouth is too small to fit 5 000 M&Ms, so the right answer must be A.
			I guessed.
			Other.
Noa: I can lift a dog that weighs 5 000 g. ***	68,9 %	36,7 %	5 000 g is 5 kg and it is possible to lift that much.
		28,9 %	5 000 textbooks are too many to fit into one bag and the mouth is too small to fit 5 000 M&Ms, so the right answer must be B. ###
		3,3 %	I guessed.
			Other.
Sofi: I can fit 5 000 M&Ms into my mouth.	14,4 %	3,3 %	If Sofi likes M&Ms, she can put 5 000 of them into her mouth at once.
		7,8 %	5 000 textbooks cannot fit into a schoolbag and 5 000 g is too heavy to lift, so the correct answer must be C.
		2,2 %	I guessed.
		1,1 %	Other.
I cannot tell.	15,6 %	11,1 %	I need more information to answer.
		1,1 %	I need to make calculations with a pen and paper to answer.
		1,1 %	I guessed.
		2,2 %	Other.

***Correct answer
NS-based method

A large majority of 68,9% answered the question correctly, relying mostly on a rule-based (36,7%) method of conversion across different measurement units (g to kg) or on a well-developed number-sense (28,9%) of intuitively understanding the magnitude of large numbers. These are better results than in the Yang (2019) findings, where only 38,1% of the pupils answered this question correctly. Over 30% of the pupils could not answer the question correctly, indicating a yet-immature understanding of magnitude in large quantities and related misconceptions such as 5 000 g being too heavy to lift (7,8%), or it being possible for 5 000 M&Ms to fit into the mouth (3,3%). This question was the one that got more answers in the category “I need more information to answer”, showing that it is still difficult for fourth graders to assess how reasonable and meaningful large numbers are in real-life situations.

4.1.2 Pupils' responses to test question 2

Table 2 Sampled pupils' responses to question 2 on the first- and second-tier tests

2. If I sum up two 3-digit numbers, how many digits will the answer have?		
3 digits	3,3 %	Because both numbers will have three digits, the answer must have three digits.
13,3 %	7,8 %	Because if I sum up for example two 3-digit numbers like 100 + 100 =200 and 200 has three digits.
	2,2 %	I guessed.
		Other
4 digits		If I sum up two 3-digit numbers, the result will be a bigger number, so it will have four digits.
1,1 %	1,1 %	If I sum up for example two 3-digit number like 500 + 500 = 1 000 and 1 000 has four digits.
		I guessed.
		Other
3 digits or 4 digits***	7,8 %	If the 3-digit numbers are small, the answer will not reach 1 000, so it will have still 3 digits. But it they are big, the answer will be 1 000 or more, so it will have 4 digits.###
22,2 %	13,3 %	If I add upp 100 + 100 = 200 and 200 has 3 digits. But if I add upp bigger 3-digit numbers like 500 + 500 = 1 000 and 1 000 has 4 digits.
	1,1 %	I guessed.
		Other
6 digits	46,7 %	Because both numbers have 3 digits and 3 + 3 = 6, the answer will have 6 digits.
55,6 %	7,8 %	I guessed.
	1,1 %	Other
I cannot tell	1,1 %	I do not remember what the word "digit" means.
7,8 %	1,1 %	I need to know what the numbers are to answer the question.
	1,1 %	I need to calculate with pen and paper to answer the question.
	2,2 %	I guessed.
	2,2 %	Other

***Correct answer

NS-based method

Table 2 summarizes pupils' answers to question 2 and their reasoning methods when giving these answers. The question relies on pupils' understanding of the concepts of "number" and "digit", which is emphasized in the mathematics content area C2 for grades 1-2 in the national core curriculum (2014, p. 138): “It is ensured that the pupils are familiar with the connection between numbers, numerals, and digits.” Additionally, it requires them to make generalizations of the addition operation based on the recognition of its properties.

Only 22,2% of the pupils answered the question correctly. The overwhelming majority of 55,6% answered that the sum of two 3-digit numbers is a 6-digit number, mostly because “the sum 3 + 3 = 6” (46,7%). Similarly, in the Yang (2019) findings, about 50% of the respondents chose the “6-digit number” alternative. While such a misconception is most likely steaming from a challenge in understanding the concepts of “number” and “digit”, only 1% of the pupils seemed aware of this challenge and answered “I do not remember what the word “digit” means”. About 21% of the pupils relied on a rule-based method to answer the question, trying to derive the answer from their experience of operations with 3-digit numbers, such as adding “100 + 100” or “500 + 500”. About 8% of the pupils relied on the number-sense method, using their intuitive understanding that the sum of smaller 3-digit numbers would remain under 1000, whereas for bigger numbers the sum would reach or exceed 1000, becoming a 4-digit number. Such results indicate a tendency among such age-group to overgeneralize based on their experience and a challenge in drawing meaningful conclusions based on their observations. Additionally, the answers reflect a need for using more often mathematics terminology and concepts (i.e. “digit”) in daily teaching to ensure the pupils properly understand and remember them.

4.1.3 Pupils’ responses to test question 3

Table 3 shows the answers to the first- and second-tier questions for item 3. The question assesses the pupils’ ability to estimate the magnitude of a measured object, as well as consider whether a measurement result is reasonable (Finnish National Board of Education, 2014, O12, p. 253). 70% of the pupils answered the question correctly, showing a good understanding of lengths and distances that are well represented in their daily-lives, such as the height of a person. In the Yang (2019) study the percentage of correct answers for this question was lower at about 46%. About 26% of the pupils opted for the alternative “260 cm” relying either on the misconception that 170 cm is too small of a number to indicate the “most likely” height of a person (4,4%), or that the height of a person grows linearly (18,9%) or. In Yang’s study (2019) this later misconception of linear growth was much higher at about 40%.

Table 3 Sampled pupils' responses to question 3 on the first- and second-tier tests

3. Emil is 10 years old and 130 cm tall. How tall will Emil most likely be when he is 20 years old?

170 cm***	20,0 %	It is possible for a person to be 170 cm tall.###
70,0 %	46,7 %	260 cm and 350 cm are too tall, so the correct answer must be A.###
	3,3 %	I guessed.
		Other
260 cm	18,9 %	Because Emil will be two times older (2 x 10 = 20) he will also be 2 times taller (2 x 130 cm = 260 cm).
25,6 %	4,4 %	170 cm is too short and 350 cm is too tall, so the correct answer must be B.
	2,2 %	I guessed.
		Other
350 cm		It is possible for a person to be 350 cm tall.
2,2 %	1,1 %	170 cm and 260 cm are too short, so the correct answer must be C.
	1,1 %	I guessed.
		Other
I cannot tell.	2,2 %	One cannot predict a person's height.
2,2 %		I need to make calculations with a pen and paper to answer.
		I guessed.
		Other

***Correct answer

NS-based method

4.1.4 Pupils’ responses to test question 4

Table 4 shows the sampled pupils’ responses to question 4, which assesses skills in understanding the reasonableness of quantities pertaining to volume measures in daily-life situations. An overwhelming majority of 91% of the pupils answered the question correctly. Most of them (46,7%) relied on the rule-based method of converting deciliters to liters and 38,9% used their estimation of 5 l as too large of an amount of milk to drink each morning. This question was modified from the original to remove the milliliter unit, which had not been covered yet in the fourth-graders teaching at the time of the test. In Yang (2019) the percentage of correct answers was 52% and about 40% of the sampled pupils believed that it was possible to drink 500 dl or 500 l of milk every morning, showing a still poor estimation of these measurement units.

Table 4 Sampled pupils' responses to question 4 on the first- and second-tier tests

4. Minna says "I can drink 5_____ of milk every morning." Which of the following measuring units can best complete the sentence?

5 dl***	46,7 %	5 dl is half a liter, so half a milk carton. It is possible to drink that much milk per day.
91,1 %	38,9 %	5 l is too large of an amount. So 5 dl must be the correct answer. ###
	3,3 %	I guessed.
	2,2 %	Other
5 l	1,1 %	Milk is usually measured in liters, so 5 l must be the correct
4,4 %	2,2 %	Dl is too small of a unit, so 5 l must be the correct answer.
		I guessed.
	1,1 %	Other
I cannot tell.	3,3 %	It depends on the person how much milk per day they drink.
4,4 %	1,1 %	I need to make calculations with pen and paper to find the
		I guessed.
		Other

***Correct answer

NS-based method

4.1.5 Pupils' responses to test question 5

Table 5 summarizes the answers to question 5. Similarly to questions 3 and 4, this question also measures skills emphasized in learning objective 12 (012) of the Finnish national core curriculum (2014, p. 253), according to which pupils are guided to estimate measurement magnitudes and evaluate the reasonableness of measuring results. Differently from question 3, this question relies also on the pupils' ability to convert among different measurement units, a skill mentioned in content area C4 of the curriculum (Finnish National Board of Education, 2014, p. 254): "[the pupils] practice making unit conversions with the most common units of measurement." The unit of millimeters was removed from the alternatives, due to pupils not being familiar with it yet. Over 83% of the pupils answered correctly (in Yang (2019) study this was about 53%), relying mostly (51,1%) on the rule-based method of converting centimeters to meters and then the number-sense method of estimating how high 3 m or 300 m is.

Table 5 Sampled pupils' responses to question 5 on the first- and second-tier tests

5. Which of the following is most likely the height from the floor to the ceiling of a classroom?		
300 cm***	51,1 %	300 cm is 3 m, and the height from floor to ceiling of a classroom can be 3 m.
83,3 %	26,7 %	300 m is too high for the height of a classroom. So 300 cm must be the correct answer.###
	4,4 %	I guessed.
	1,1 %	Other
300 m		The height of a classroom is usually measured in meters, so 300 m is the correct answer.
12,2 %	6,7 %	300 cm are too small, so 300 m is the correct answer.
	4,4 %	I guessed.
	1,1 %	Other
I cannot tell.	2,2 %	I need to see the classroom to find out the answer.
4,4 %		I need to make calculations with pen and paper to find the answer.
	1,1 %	I guessed.
	1,1 %	Other

***Correct answer

NS-based method

Compared to question 3, the pupils had less misconceptions in this question, indicating their number sense is better developed if the difference between the quantities is large (there is a bigger difference between 3 m and 300 m, compared to between 170 cm and 260 cm). Compared to question 4, the percentage of wrong answers in this question was higher, indicating possibly that daily-life experience enhances the development of number-sense (pupils have more experience with milk cartons in their daily life than with estimating or measuring room heights).

4.1.6 Pupils' responses to test question 6

Findings for question 6 are summarized in table 6. The question measures a pupil's mental arithmetic skills, making use of properties of operations (Finnish National Board of Education, 2014, O 10, p. 253), as well as his or her abilities in assessing whether a given solution to a computational problem is reasonable (Finnish National Board of Education, 2014, O 6, p. 253).

Table 6 Sampled pupils' responses to question 6 on the first- and second-tier tests		
6. Which of the results is the best answer for 149 x 4= ______?		
Les than 500	7,8 %	149 is close to 100 and 100 x 4 = 400, which is less than 500.
14,4 %	2,2 %	600 and 700 are too big, so the answer must be less than 500.
	3,3 %	I guessed.
	1,1 %	Other
Between 500 and 600***	24,4 %	149 is a little less than 150, and 150 x 4 = 600, so the answer will be a little less than 600.###
58,9 %	30,0 %	I calculated that 149 x 4 = 596, which is between 500 and 600.
	4,4 %	I guessed.
		Other
Between 600 and 700	15,6 %	149 is close to 150 and 150 x 4 = 600, so the answer will be between 600 and 700.
21,1 %	4,4 %	I guessed.
	1,1 %	Other
I cannot tell.	4,4 %	I need to make calculations with pen and paper to find the answer.
5,6 %		I guessed.
	1,1 %	Other

***Correct answer

NS-based method

About 59% of the pupils answered the question correctly, relying mostly on the rule-based method (30%) of computing the result and on the number-sense method (24,4%) of rounding and making approximate estimations. From the 41% of the pupils that gave the wrong answer, the majority still relied on rounding and estimating, but displayed misconceptions either in the rounding process (7,8%) or in what direction to approximate once the rounding has been done correctly (15,6%). In yang (2019) study the results were very similar: 59% answered correctly and misconceptions in rounding and approximating directions accounted for most of the incorrect answers. The answers reflect a need for more practice with estimating the range of the computational result before such exact computation is performed.

4.1.7 Pupils' responses to test question 7

Sampled answers for question 7 are summarized in table 7. Similarly to question 6, this question also measures the ability to mentally perform arithmetical operations, while relying on number and operations' properties.

Table 7 Sampled pupils' responses to question 7 on the first- and second-tier tests

7. Which number is the best choice, if you want to get the highest answer from the computation "458 - ____"?		
202***	76,7 %	To get the highest answer I need to subtract the lowest number and 202 is smaller than 210 and 258.###
83,3 %	4,4 %	I calculated all the answers. I got the highest answer when I subtracted 202.
	2,2 %	I guessed.
		Other
210	3,3 %	I calculated all the answers. I got the highest answer when I subtracted 210.
3,3 %		I guessed.
		Other
258	1,1 %	258 is bigger than 202 and 210, so the highest answer is when I subtract 258.
6,7 %	2,2 %	I calculated all the answers. I got the highest answer when I subtracted 258.
	2,2 %	I guessed.
	1,1 %	Other
I cannot tell.	3,3 %	I need to make calculations with pen and paper to find the answer.
6,7 %	2,2 %	I guessed.
	1,1 %	Other

***Correct answer

NS-based method

An overwhelming majority of over 83% of the pupils answered correctly, which was a higher percentage than in the Yang (2019) study where the number of correct answers was about 60%. From the correct answers, the majority (76,7%) relied on the number-sense method by judging that when the subtracted number is the lowest, the result is the highest. Other method to solve the problem have included the rule-based method of trying to perform all the computations mentally, or guessing. About 3% of the pupils have said that they would need to make pen-and-paper calculations to be able to answer. In question 6 such answer was given by about 4% of the respondents. While these percentages are quite low, it shows that not all fourth graders have reached confidence in mental arithmetics yet.

4.1.8 Pupils' responses to test question 8

Table 8 Sampled pupils' responses to question 8 on the first- and second-tier tests

8. Food A costs 50 euros for 1,5kg and food B costs 100 euros for 3,5kg. Which food is cheaper?

Food A	28,9 %	50 euros is less than 100 euros, so food A is cheaper.
44,4 %	11,1 %	I divided 50 by 1,5 and 100 by 3,5 to get the price of 1 kg for each food. I found that food A is cheaper.
	2,2 %	I guessed.
	2,2 %	Other
Food B***	20,0 %	I divided 50 by 1,5 and 100 by 3,5 to get the price of 1 kg for each food. I found that food B is cheaper.
48,9 %	21,1 %	With 100 euros I can buy 3 kg food A and 3,5 kg of food B, so food B is cheaper.###
	4,4 %	I guessed.
	3,3 %	Other
Equally cheap	1,1 %	I divided 50 by 1,5 and 100 by 3,5 to get the price of 1 kg for each food. I found that food A and B cost the same.
3,3 %	1,1 %	I guessed.
	1,1 %	Other
I cannot tell.	2,2 %	I need to know the price for 1 kg for each food to compare.
3,3 %	1,1 %	I need to make calculations with pen and paper to find the answer.
		I guessed.

***Correct answer

NS-based method

Table 8 shows the answers given to question 8. The question involves the challenge of multi-step problem of first identifying the numbers that pertain to the same category (50 euros and 1,5 kg for category A and 100 euros and 3,5 kg for category B); second identifying the correct operation to be performed with these numbers (division of quantity by price); third connecting the result with the right category; and fourth interpreting the meaning of the result (the smaller number of euros per kg indicates the cheaper food). Without pen-and-paper calculations the steps naturally become too many and it is likely that mistakes happen along the way. Such mistakes could be avoided in the number-sense method where instead of performing all the above-mentioned computational steps mentally, the pupils identify the amount of 100 euros as a benchmark and realize that with that same amount, one could purchase 3 kg of food A but 3,5 kg of food B. Since a larger quantity of food B can be purchased with the same amount of money, food B is cheaper. Even with the benchmark in mind, the pupils would still need to remember which numbers went to what category and perform mental arithmetics (i.e. 100 / 50 = 2 and 2 * 1,5 = 3 and 3 < 3,5).

The challenge of the questions is reflected also in the answers. About 49% of the pupils answered correctly, relying mostly on the number-sense method (about 24%, since the answers classified under “other” reflected mostly a number-sense approach as well). About 32% of the pupils correctly identified the rule-based strategy for solving this problem: dividing the amount of money by the food quantity to find the price per kilo. Most of them (20% of the overall respondents) were able to perform all the computational steps correctly and derive the correct answer, whereas 11% came to the wrong conclusion, despite the implementation of a correct strategy. The most common misconception identified in this question (about 29%) is that pupils may neglect the food quantities and compare only the amount of money when comparing the prices. This was the most common misconception held by the respondents in the Yang (2019) study as well, although in this study this question was answered correctly by a majority of about 61% of the pupils.

4.1.9 Pupils’ responses to test question 9

Question 9 (answers summarized in table 9) assesses pupils’ ability to perform mental computations with fractions. This question was not included in the Yang (2019) study, but it was added to this study considering the important part fractions and fraction operations have in the fourth-grade curriculum in Finland. About 38% of the pupils answered the question correctly. The most commonly-held misconceptions were those of considering only the numerators and neglecting the enumerators in fraction operations (25,5%) or thinking that fractions only represent quantities that are smaller than 1 (4,4%). Additionally, this was the question where the answer “I cannot tell” had the highest percentage (20%). Part of the difficulty in answering this question may have resulted from fraction operations being still relatively new to fourth-graders and as a result the pupils have not yet gained fluency in mental operations in this arithmetic area.

Table 9 Sampled pupils' responses to question 9 on the first- and second-tier tests

9. Which of the following is the best answer for 17/19 + 2/19 = _____?

Less than 1	4,4 %	The sum of two fractions is a fraction and fractions are smaller than 1.
10,0 %	4,4 %	I calculated 17/19 + 2/19 and the answer I get is less than 1.
	1,1 %	I guessed.
		Other
Equal to 1***	1,1 %	17/19 is less than 1 and 2/19 is less then 1, so the sum will be equal to 1.
37,8 %	33,3 %	17 + 2 = 19 and 19/19 is equal to 1.###
	3,3 %	I guessed.
		Other
More than 1	14,4 %	17 and 2 are bigger than 1, so the sum will also be bigger than 1.
32,2 %	11,1 %	17 + 2 = 19 and 19 is bigger than 1.
	6,7 %	I guessed.
		Other
I cannot tell.	11,1 %	I need to make calculations with pen and paper to find the answer.
20,0 %	2,2 %	I guessed.
	6,7 %	Other

***Correct answer
NS-based method

Another explaining factor might be the notation sign that was used for fractions in the test "/". In 3 out of the four participating classes the pupils were seeing such a notation for the first time. The notation was explained before the pupils started the test along with the other instructions and pupils could consult the white board in front of the classroom during the test, where it was written that:

$$17/19 \longrightarrow \frac{17}{19} \text{ and } 2/19 \longrightarrow \frac{2}{19}.$$

However, pupils might need more time to internalize new notations without increased challenge in solving problems that contain such notations.

4.1.10 Pupils' responses to test question 10

Table 10 shows the sample answers to question 10, which assesses pupils' understanding of the properties of decimal number operations. Similarly to question 9, this question was not part of the Yang (2019) test, but was included in this study for the same reason as before.

Table 10 Sampled pupils' responses to question 10 on the first- and second-tier tests

10. If I sum up two numbers that have one decimal each, how many decimals will the answer have?

Zero decimal		2,2 %	When I sum up for example 0,5 + 0,5 = 1 and the answer has 0 decimals.
	4,4 %	1,1 %	I guessed.
		1,1 %	Other
One decimal		3,3 %	If both numbers have one decimal the answer will also have one decimal.
	13,3 %	6,7 %	If I sum up for example 0,1 + 0,1= 0,2 and 0,2 has one decimal.
		3,3 %	I guessed.
			Other
Zero decimals or one decimal***		4,4 %	If I sum up for example 0,5 + 0,5 = 1 so the answer has 0 decimals, but if I sum up 0,1 + 0,1= 0,2 so the answer has one decimal.
	6,7 %	2,2 %	If the decimals sum is smaller or bigger than 10, the answer will have one decimal. But if the decimals sum is 10, the answer has 0 decimals.###
			I guessed.
			Other
Two decimals		44,4 %	One decimal from one number + one decimal from the other number, the answer will have two decimals.
	61,1 %	10,0 %	If I sum up for example 0,5 + 0,6 = 0,11 so the answer has two decimals.
		5,6 %	I guessed.
		1,1 %	Other
I cannot tell		11,1 %	I do not remember what the word “deciimal” means.
	14,4 %	1,1 %	I need to make calculations with pen and paper to find the answer.
		1,1 %	I need to know what the numbers are to answer.
		1,1 %	I guessed.
			Other

***Correct answer
NS-based method

This is the question that had the lowest percentage of correct answers (6,7%), out of which 2% relied on the number-sense method of generalizing the decimal property that when their sum is larger than 10, they give one whole unit plus the remaining part expressed by one decimal.

About 4% could correctly derive the right answer based on their experience with decimal number operations. About 10% used this same strategy of trying to conclude based on familiar operations with decimals, but it was challenging for them to take into consideration all the different scenarios—7% neglected cases where the decimals' sum would be 10 focusing only on cases such as 0,1 + 0,1 = 0,2; whereas 2% neglected such cases and considered only scenarios similar to 0,5 + 0,5 = 1. The overwhelming misconception (61%) however seems to be the belief that the sum of two numbers with one decimal each, is a number with two decimals because "one decimal + one decimal = two decimals" (44%) or because "0,5 + 0,6 = 0,11" (10%).

Such a misconception might be explained at least partially with challenges with the terminology, as shown also by the high percentage (11%) of the pupils that are aware of not remembering what the concept "decimal" means. Another explaining factor, could be the fact that fourth graders are still quite new to operations with decimal numbers. Additionally, at this age pupils might find it challenging to correctly generalize their knowledge in the form of a rule, although they would probably be able to correctly perform operations in a certain arithmetic area. More experience in the area, as well as higher exposure to such questions that seek for such generalizing rules could potentially help pupils to gain better understanding and higher accuracy in judging the reasonableness of computational results.

4.2 Pupils' overall performance in judging the reasonableness of a computational result and related misconceptions

Table 11 summarizes the overall pupil performance for all the first- and second-tier questions. For the first-tier test, the percentage of the correct answers is reported for each question. For the second-tier test, pupils' problem-solving strategies have been classified based on the method/reasoning employed by them into four main categories of number-sense method (NS), rule-based method (R), misconceptions (M), and guessing (G) and the corresponding percentages have been reported for each category. In addition, percentages of other categories were reported for the questions where they are the most relevant, where pupils thought they could not answer the question out of a need for more information (MI), pen-and-paper calculations (PC), or explanation of the terminology used (T). The process and reasoning behind the forming of these categories is explained under the Data Analysis section. Last the table reports for which questions the chi-squared test results were significant at 1% significance value (p=0,01).

Table 11 Pupil performance in judging the reasonableness of a computational result

	First-tier test	Second-tier test (%)								
Question	Correct (%)	NS	R	M	G	MI	PC	T	O	Chi-square significance
1	68,9	28,9	**36,7**	13,3	6,7	14,4	-	-	-	***
2	22,2	7,8	23,3	**50,0**	12,2	-	-	-	6,7	***
3	70,0	**66,7**	-	26,7	6,7	-	-	-	-	***
4	91,1	38,9	**46,7**	8,9	5,6	-	-	-	-	
5	83,3	26,7	**51,1**	14,4	7,8	-	-	-	-	***
6	58,9	24,4	**31,1**	30,0	14,4	-	-	-	-	***
7	83,3	**76,7**	10,0	5,6	7,8	-	-	-	-	***
8	48,9	27,8	31,1	**33,3**	7,8	-	-	-	-	***
9	37,8	32,2	5,6	**37,8**	13,3	-	11,1	-	-	***
10	6,7	-	13,6	**63,6**	11,4	-	-	11,4	-	***
Average for all the items		36,7	27,7	28,4	9,4					

Bold numbers indicate the most frequent category of answers for the second-tier questions
*** - Chi-square test significant at p=0.01 significance level
NS - number-sense based method
R - rule-based method
M - misconceptions
G - guessing
MI - need for more information
PC - need for paper calculations
T - challenges in understanding the terminology used in the question
O - sum of MI, PC, T categories when all are less than 5 and of almost equal frequency

The average correct answer rate for all the ten items varied between 6,7% - 91,1%, with the total average for all the items being 57,11%. For the first-tier questions 1-8 that were based on the Yang (2019) instrument, the total average of correct answers was 65,83%. In Yang's (2019) findings this percentage was 51,7%.

For 6 out of 10 questions the majority of the pupils (over 50%) answered correctly. The questions that were answered correctly only by a minority of pupils (below 50%) mainly asked for a generalizing conclusion based on the properties of arithmetic operations with multiple-digit numbers (question 2) and decimal-numbers (question 10); or involved fraction computations (question 9) or multiple-step mental operations (question 8). These results indicate a need for more practice in these areas and with these types of questions that advance a pupil's number-sense by asking them to estimate and use their judgment, rather than to simply compute and report the results.

The chi-squared statistics were meaningful at 1% significance level for all questions, except for question 4. This means that in question 4 (estimation of reasonable volume quantities) the distribution of pupils' reasoning among the categories of NS, R, M and G are not significantly different from what could be explained by random chance (which in this study was assumed to be an equal distribution across all categories). We can say that pupils rely almost equally on their number-sense as they do on rules when answering this question, whereas misconceptions and guessing are both represented in low frequencies that do not seem different in statistical terms. For this reason, question 4 is not referred to in the analysis following in the paragraph below. In all other questions though, the chi-squared tests are statistically significant, showing the frequencies of answers in all the represented categories cannot be explained by chance, but indeed pupils tend to rely more on certain problem-solving strategies over the rest when answering these questions.

The number-sense method was not meaningfully represented in generalizing operation properties for decimal numbers (question 10) and had a low percentage of about 8% when generalizing rules of operations with multi-digit numbers (question 2). In other questions, a significant amount of pupils relied on a well-developed number-sense to solve problems, especially in questions 3 (estimating a person's height) and 7 (arithmetic deduction), where the number-sense method prevailed over all others and was used by over half of the responding pupils (67% and 77% of the pupils respectively).

The rule-based method followed by misconceptions seem to explain most of the answers across all the 10 questions. The rule-based method prevails in questions pertaining to the estimation of quantities (question 1), volume (question 4), distances (question 5), and estimating the range of multiplication operations (question 6). Misconceptions on the other hand are the most represented category in questions generalizing number operations (questions 2 and 10), containing fractions (question 9), or involving multi-step mental arithmetics (question 8). Guessing and other categories seem to not have had the highest frequency for any of the questions asked.

In sum, these results show that fourth graders rely both on memorized rules and exact computations, as well as on number-sense strategies in problem-solving. Yet their number-sense needs to be further practiced and developed in most areas of mathematics. Pupils show also different misconceptions when employing their judgement to assess the reasonableness of a computational result without using pen and paper calculations. A broader discussion of these findings and their relation to prior research follows in the next section.

5 Discussion and conclusion

The aim of this study was (i) to measure the performance of Finnish fourth graders in judging the reasonableness of a computational result in mathematics, as well as (ii) to identify some of the related misconceptions that these pupils rely upon when judging this reasonableness. This section provides answers to these two research questions based on the results obtained from the research data analysis and relates these results to prior research findings in the field.

Fourth graders perform above average in identifying unreasonable answers to mathematical problems

The performance of 90 Finnish fourth-graders in judging the reasonableness of computational results reached a total average score across all the ten questions of 57,11%. This shows fourth graders perform above average in identifying unreasonable answers to mathematical problems. Although this test did not measure pupil abilities in mathematical computations and there are no national standardized tests in Finland to compare these scores to, teachers of the participating classes commented the results to be lower than what the pupils would typically achieve in a traditional mathematics test pertaining to pen-and-paper calculations. This is in line with prior research that has shown pupils' performance in calculations to be significantly higher than that in number-sense related questions (Reys & Yang, 1998; Bana & Dolma, 2006; Mohamed & Johnny, 2010).

If we count only the eight items found in equivalent international instruments the percentage of correct answers in total is 65,83%, whereas in the Yang (2019) study it was reported to be 51,7%. However, the exact performance scores of Finnish fourth graders should not be compared directly to those recorded in international studies, since there have been differences in the sample size and diversity, the instrument, and the conditions in which the pupils took the test. To illustrate, the Yang (2019) study included 790 Taiwanese fourth graders across a variety of schools; the instrument was composed of 8 items (as opposed to 10 questions used in the Finnish setting); and the test on judging results' reasonableness constituted only one of the four parts that the overall test given to pupils had. Moreover, no statistical tests for difference have been performed to compare the Finnish and international scores. Keeping these limitations in mind, we could still note that the sampled Finnish pupils seem to perform slightly better at identifying reasonable answers to mathematical problems, than their international peers that have participated in similar studies (McIntosh et al., 1997; Yang, 2019).

This difference could be explained with the mathematics instruction in Finland relying heavily on the use of physical manipulatives, which research has shown to advance the development of number sense (Gurganus, 2004; Witzel et al., 2012). Another explaining factor could be the inclusion of the ability to judge whether a result is a reasonable and meaningful answer to a problem among the mathematics learning objectives in the national Finnish curriculum (Finnish National Board of Education, 2014). International research has attributed low performance in assessing reasonableness of answers to the fact that such an area seems to be missing all together from mathematics learning goals set in the national curricula (Alajmi & Reys, 2007; Li, Yang & Lin, 2015).

Fourth graders have challenges with fractions and decimals

A comparison of the percentage of correct answers across the different questions, shows pupils to have a better understanding of the effect of operations on whole numbers, whereas the areas of fractions and decimals (questions 9 and 10) pose still challenges for the fourth graders. This is not a surprising finding, given that in the development model of number sense, it is suggested that by the age of 10 children have typically reached a good general and specific understanding of the properties of numbers and mathematical operations in the whole numbers domain, but in rational numbers this understanding is still at lower levels for this age group (Kalchman, Moss & Case, 2001).

Fourth graders reveal a variety of misconceptions when judging the reasonableness of computational results

An analysis of the pupils' answers to the second-tier questions of the test shows that fourth graders reveal a variety of misconceptions when judging the reasonableness of computational results. Prior research has also shown that learner misconceptions are both numerous and diverse and both their identification and especially their correction poses challenges to educators (Novak, 2002; Allen, 2014). All the four main types of mathematics misconceptions as identified in the Ryan & Williams (2007) taxonomy seem to be present among the misconceptions revealed by this study. A short elaboration on each type follows below.

Modelling refers to the ability to meaningfully connect mathematics to the real world (Ryan & Williams, 2007), which is very important when judging the reasonableness of answers, since often such judgment means pondering on whether this result is reasonable or not in a real world scenario (Alajmi & Reys, 2007). Modelling misconceptions refer to situations where the magnitude of numbers or properties of operations have not been linked correctly to a real world model (Ryan & Williams, 2007). Examples of modelling misconceptions revealed in this study were the belief that 5000 M&M candies can fit at once into a person's mouth (14,4%) in question 1; thinking that a person's height grows linearly as a function of age (18,9%) in question 3; or estimating the height of a classroom to be 300m (12,2%) in question 5. Such misconceptions seemed to prevail also in the Yang (2019) study, showing the challenges this group age faces in correctly modelling mathematical problems.

Prototyping misconceptions on the other hand, result from the learner holding on to a typical example of the concept, even when that concept is mathematically incorrect (Ryan & Williams, 2007). For instance, the learner associates the concept "cheaper" with "less money", which then may lead him or her to draw wrong conclusions (44,4%) when asked to compare which of the two types of food is cheaper, when the cost is reported for different quantities of each food (question 8). Similarly, children might hold on to a concept of fractions indicating a quantity that is smaller than 1 (4,4%) as shown in question 9, if most of the fractions they have encountered have been of such a type.

Misconceptions related to the overgeneralization of whole number properties into the rational numbers domain were noticed in this study mostly in question 9, where pupils estimated the sum of 17/19 and 2/19 to be larger than 1, because 17 and 2 are larger than 1 (32,2%). Whereas overgeneralizations count for most of the computational mistakes made by pupils 8-16 years old (Ryan & Williams, 2007), they do not seem to represent the most frequent type of misconceptions in this study. Part of the reason might be that this study does not directly measure computational skills. Another explaining factor could be the low amount of questions involving rational numbers, where most of the overgeneralization mistakes happen as a result of children drawing on whole number properties.

The last type of misconceptions steams from challenges in process-object linking, where the learner does not understand through what mathematical operations (i.e. process) he or she can derive the answer to a mathematical problem (i.e. object) (Ryan & Williams, 2007). In this test, the most typical examples of this misconception type can be noticed in questions 2 and 10, where 46,7% of the pupils thought that to find out the amount of digits/decimals in the sum, they must sum up the amount of digits/decimals in each of the addends. Due to this process-object linking misconception, they concluded that the sum of two 3-digit numbers is a 6-digit number and the sum of two addends with one decimal each must have two decimals.

The relatively high frequency of misconceptions between 5,6% - 63,6%, with a total average for all the ten items being 28,4% shows that on average, over a quarter of the pupils rely on wrongly constructed mathematical concepts to judge how a reasonable or meaningful an answer is. This indicates a need for learning materials and instruction to focus even more on the Finnish national curriculum goal of guiding the pupil "to develop his or her skills in assessing whether the solution is reasonable and meaningful" (Finnish National Board of Education, 2014, p. 252).

Implications for teaching

The importance of the learning environment and teaching in the development of the number sense among children has been continually emphasized in research (Deheane, 2011; Berch, 2005). Prior research has criticized the focus of mathematics textbooks on algorithmic computations, whereas exercises asking pupils to judge how reasonable an answer is to a certain problem seem to be missing all together (Yang, 2019). There is no research assessing Finnish mathematics textbooks in this aspect, but from the researcher's personal experience, it may be concluded that they could be criticized on this same bases. There seems to be a need therefore, for the developers of textbooks and other learning materials to become better aware of the different components of number sense, as well as strategies to develop them among children in basic education.

Educators have a crucial role in this. It is important that teachers are familiar with the concept of number sense, its importance, how it can be developed, and how it could be measured. Research has expressed concern on the lack of awareness teachers seem to have regarding the gap between what they aim to promote among learners—meaningful understanding of conceptual knowledge—and what their practices seem to develop instead—mechanical application of learned rules (McIntosh et al., 1997). Pre-service and in-service teacher training needs to focus on enabling teachers in noticing such a gap and addressing it.

While identifying and especially correcting misconceptions emerging in mathematics learning is a considerable challenge for all educators (Novak, 2002; Allen, 2014), there is a need for all teachers to become aware of strategies how pupils' misconceptions could be identified, classified into frequent types, and addressed in teaching. Ryan & Williams (2007) for example, urge for mathematics teachers to employ a number of strategies to replace mathematics misconceptions with correctly constructed concepts, such as: (i) continuous use of physical manipulatives and real-life scenarios to help children properly model mathematical operations; (ii) exposing children to a variety of examples that aim to challenge the most typical example of a concept that the pupil might have in mind, so that mathematically concepts replace prototypical concepts (for instance, examples where fractions represent amounts greater than 1, etc.); (iii) asking pupils questions so that they make their (over)generalizations explicit and then asking further for examples when such generalizations do not hold; and (iv) introducing different models and continuously engaging pupils to see the links between these models, so that they become fluent in process-object linking in mathematics. The constructivist theories of Piaget and Vygotsky on challenging learners at the zone of their proximal development, enabling them to solve the presented cognitive conflicts by correcting their misconceptions are of paramount relevancy and importance in this process.

Taking upon the challenge of developing higher quality instruction and learning materials that support the development of number sense in children will lead to more meaningful learning and higher mathematical proficiency for all pupils.

6 Reliability of the study and recommendations for further research

Much caution should be exercised when generalizing the findings of this study. The sample size of 90 fourth graders is relatively small and it is drawn from only one school. Although adapted from high-quality published international research, the test instrument has its own limitations in fully capturing pupils' ability to identify reasonable answers, as number sense components are complex and exhibit themselves in a multifaceted array of mathematical skills. The designing of the second-tier questions to capture the reasoning of pupils when judging on the reasonableness of results posed challenges, as it might be difficult to predict all the possible strategies pupils are likely to employ in their solving of mathematical problems. The data collection method has also its own limitations. As one of the teachers of the participating classes noted, pupils would probably be more able to answer correctly to some of the questions, if they were orally asked—in the written form however, such questions might be more challenging to comprehend for some. Some pitfalls inherent to the methodology employed in the data analysis are also worth-mentioning—for instance the chi-square test gives no means for drawing the different categories into which the data is classified, or for assessing whether the used categories are correct. As a result, when classifying pupils' answers from the second-tier test, a strategy that may have been classified as a "rule-based method" could also display a well-developed number sense.

While the above-listed limitations of the study call for a critical interpretation of the results—as should be the case in all scientific research—a number of steps were taken to ensure the validity of the study. First, the sample size of 90 pupils, although small, allows for statistical inference of the employed methods (Lindsay, 1995). Prior studies have concluded Finland to have "a very small variance in student performance across schools" (Sahlberg, 2007, p. 158), which might lead us conclude that performance in other schools probably would not vary significantly. The instrument items, along with the second-tier questions to capture pupils' reasoning, and the categories for classifying their solution strategies have been already tested and reported as effective in studying performance in judging the reasonableness of computational results and related misconceptions, by studies published in good-quality international journals (Lin, Yang, & Li, 2015; Yang, 2019).

Although the study is quantitative in nature, the interpretation of the obtained results is not immune to researcher subjectivity and some degree of academic bias. Because 80% of the test items were adopted from an international instrument, it was tempting to report the obtained results side by side to those reported by Yang (2019). However, this side-by-side reporting should not be generalized to an exact comparison, given the differences in sample size and diversity and in the data collection setting. Moreover, no statistical methods were employed to test whether the results obtained here are significantly different from those reported by Yang (2019). Nevertheless, reporting the Yang (2019) findings alongside this study's results adds to our meaningful understanding of these findings. It is of interest, for instance, that children of the same age reveal similar misconceptions and moreover to a similar frequency, although their upbringing and formal instruction environments have been so different. It shows something about the universal nature and power of misconceptions in learning.

Subjectivity is to be found not only in the interpretation of the findings, but also in the tone through which they are reported. While only slightly lower results were considered in international research to be "unsatisfactory performance", or "lack of ability" (Bana & Dolma, 2004; Alajmi & Reys, 2010; Yang, 2019), in this study not-so-different scores were labelled as "above-average", or "developing skills." Part of this difference in tone results from the format of the research—there is probably a higher need to problematize a phenomenon in a journal article than in a master's thesis monograph; part comes from different cultural and personal styles of communication. Readers are encouraged to form their own opinion on what story the reported findings tell.

Future research could start by addressing some of the limitations of this study, for instance increasing the sample size, drawing samples from numerous schools across different Finnish cities, and testing other grades in addition to the fourth grade. Another suggestion would be to use instruments that measure also the other components of number sense as identified in previous research, so that we get a broader picture on the extent to which our pupils' overall number sense is developed, not only as measured by their ability to identify reasonable answers. It would also be important to investigate Finnish teachers' views on number sense and their perceptions on the Finnish curriculum goal of developing pupils' ability to judge the reasonableness of computational results. Another study of interest would be to evaluate how well current textbooks have integrated this new curriculum addition into their learning material.

While its findings should be interpreted critically, this study shows there is a need for further research in this field so that we can contribute to improving children's mathematical proficiency across all levels of basic education.

References

Alajmi, A., & Reys, R. (2007). Reasonable and reasonableness of answers: Kuwaiti middle school teachers' perspectives. *Educational Studies in Mathematics*, 65(1), 77-94.

Alajmi, A., & Reys, R. (2010). Examining eighth grade Kuwaiti students' recognition and interpretation of reasonable answers. *International Journal of Science and Mathematics Education*, 8(1), 117-139.

Allen, M. (2014). Misconceptions in primary science (2nd ed.). Maidenhead, Berkshire: Open University Press.

Aunio, P. (2006). Number sense in young children: (inter)national group differences and an intervention programme for children with low and average performance. Helsinki: Helsingin yliopisto.

Aunio, P., Ee, J., Lim, S. E. A., Hautamäki, J., & Van Luit, J. (2004). Young children's number sense in Finland, Hong Kong and Singapore. *International Journal of Early Years Education*, 12(3), 195-216.

Aunio, P., Niemivirta, M., Hautamaki, J., Van Luit, Johannes E. H., Shi, J., & Zhang, M. (2006). Young children's number sense in china and Finland. *Scandinavian Journal of Educational Research*, 50(5), 483-502.

Ausubel, D. P. (1968). Educational psychology: A cognitive view. New York [N.Y.]: Holt, Rinehart and Winston.

Bachot, J., Gevers, W., Fias, W., & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: Orientation of the mental number line. *Psychology science*, 47(1), 172.

Baker, S., Gersten, R., Flojo, J., Katz, R., Chard, D., & Clarke, B. (2002). *Preventing mathematics difficulties in young children: Focus on effective screening of early number sense delays* (Vol. 305). Technical Report.

Bana, J., & Dolma, P. (2004). The relationship between the estimation and computation abilities of Year 7 students. *Proceedings of the 27th annual conference of the Mathematic Education Research Group of Australasia* (Vol. 1, pp. 63-70). Townsville: MERGA.

Baruk, S. (2016). Age of the captain: On errors in mathematics. Paris: Editions du Seuil.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of learning disabilities*, 38(4), 333-339.

Bobis, J. (1996). Visualisation and the development of number sense with kindergarten children. In Mulligan, J. & Mitchelmore, M. (Eds.) *Children's Number Learning: A Research*

Monograph of the Mathematics Education Group of Australasia and the Australian Association of Mathematics Teachers. Adelaide: AAMT.

Case, R. (1998). A psychological model of number sense and its development. In *Annual Meeting of the American Educational Research Association*, San Diego.

Daniels, H. (1996). An introduction to Vygotsky. London: Routledge.

Dehaene, S. (1997). The number sense: How the mind creates mathematics. New York: Oxford University Press.

Dehaene, S. (2011). The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition. Oxford University Press.

Durkin, K., & Rittle-Johnson, B. (2015). Diagnosing misconceptions: Revealing changing decimal fraction knowledge. *Learning and Instruction*, 37, 21-29.

Ekenstam, A. (1977). On children's quantitative understanding of numbers. *Educational Studies in Mathematics*, 8, 317-332.

Finnish National Board of Education[2] (2004). National core curriculum for basic education 2004: National core curriculum for basic education intended for pupils in compulsory education. Helsinki: Finnish National Board of Education.

Finnish National Board of Education (2014). National core curriculum for basic education 2014. Helsinki: Finnish National Board of Education.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of special education*, 33(1), 18-28.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of learning disabilities*, 38(4), 293-304.

Gurganus, S. (2004). Promote number sense. Intervention in School and Clinic, 40(1), 55-58. doi:http://dx.doi.org/10.1177/10534512040400010501.

Hitchcock, D. B. (2009). Yates and contingency tables: 75 years later. *Electronic Journal for History of Probability and Statistics*, 5, 1-14.

Howden, H. (1989). Teaching number sense. *The Arithmetic Teacher*, 36(6), 6. Retrieved from https://search.proquest.com/docview/208776791?accountid=13031.

Kalchman, M., Moss, J., & Case, R. (2001). Psychological models for the development of mathematical understanding: Rational numbers and functions. *Cognition and instruction: Twenty-five years of progress*, 1-38.

[2] Nowadays. Finnish National Agency for Education

Kilpatrick, J., National Research Council Staff & Swafford, J. (2002). Helping Children Learn Mathematics. National Academies Press.

Lancaster, H. O., & Seneta, E. (2005). Chi- square distribution. *Encyclopedia of biostatistics*, 2.

Lin, Y. C. , Yang, D.R., & Li, M.N. (2015). Diagnosing Students' Misconceptions in Number Sense via a Web-Based Two-Tier Test. *Eurasia Journal of Mathematics, Science & Technology Education*, 12(1).

Lindsey, J. K. (1995). Introductory statistics: A modelling approach. Oxford: Clarendon Press.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child development*, 82(4), 1224-1237.

McIntosh, A., Reys, B., Reys, R., Bana, J., & Farrell, B. (1997). Number sense in school mathematics: Student performance in four countries. Perth: Edith Cowan University.

Menon, R. (2004). Elementary school children's number sense. *International Journal for Mathematics Teaching and Learning*, 57, 1-16.

Metsämuuronen, J., & Tuohilampi, L. (2014). Changes in achievement in and attitude toward mathematics of the Finnish children from grade 0 to 9—A longitudinal study. *Journal of Educational and Developmental Psychology*.

Mohamed, M., & Johnny, J. (2010). Investigating number sense among students. *Procedia-Social and Behavioral Sciences*, 8, 317-324.

National Council of Teachers of Mathematics (Ed.). (2000). Principles and standards for school mathematics (Vol. 1). National Council of Teachers of Mathematics.

NCTM (1987). Commission on Standards for School Mathematics of the National Council of Teachers of Mathematics. *Curriculum and Evaluation Standards for School Mathematics 1987*. Reston, Va.

Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science education*, 86(4), 548-571.

Paulus, J. A. (1988). Innumeracy: Mathematical Illiteracy and its consequences. New York: Vintage Books.

Piaget, J. (1952). Child's Conception of Number. New York: Norton.

Piaget, J. (1955). The child's construction of reality. London: Routledge & Kegan Paul.

Plackett, R. L. (1983). Karl Pearson and the chi-squared test. *International Statistical Review/Revue Internationale de Statistique*, 59-72.

Reys, R. E., & Yang, D. C. (1998). Relationship between computational performance and number sense among sixth-and eighth-grade students in Taiwan. *Journal for Research in Mathematics Education*, 225-237.

Ryan, J. & Williams, J. (2007). Children's mathematics 4-15: Learning from errors and misconceptions. Maidenhead: McGraw-Hill.

Sahlberg, P. (2007). Education policies for raising student learning: The Finnish approach. *Journal of education policy*, 22(2), 147-171.

Wilson, A. J., & Dehaene, S. (2007). Number sense and developmental dyscalculia. *Human behavior, learning, and the developing brain: Atypical development*, 2, 212-237.

Witzel, B. S., Ferguson, C. J., & Mink, D. V. (2012). Number sense: Strategies for helping preschool through grade 3 children develop math skills. *Young Children*, 67(3), 89-94. Retrieved from https://search.proquest.com/docview/1140140124?accountid=13031.

Yang, D. C. (2015). Teaching and learning of number sense in Taiwan. *The first source-book on Asian research in mathematics education: China, Korea, Singapore, Japan, Malaysia and India*. Charlotte, NC: Information Age Publishing.

Yang, D. C. (2019). Performance of fourth graders when judging the reasonableness of a computational result. *International Journal of Science and Mathematics Education*, 17(1), 197-215.

Yang, D. C., & Lin, Y. C. (2015). Assessing 10-to 11-year-old children’s performance and misconceptions in number sense using a four-tier diagnostic test. *Educational Research*, 57(4), 368-388.

Yang, D. C., Li, M. N., & Lin, C. I. (2008). A study of the performance of 5th graders in number sense and its relationship to achievement in mathematics. *International Journal of Science and Mathematics Education*, 6(4), 789-807.

Appendix 1: The test instrument used in the study

1. Mikä seuraavista väittämistä on mahdollinen:
    a. Emmi: Reppuuni mahtuu 5 000 koulukirjaa
    b. Noa: Jaksan nostaa koiran, joka painaa 5 000 g
    c. Sofi: Suuhuni mahtuu 5 000 M&M -karkkia
    d. En osaa sanoa

A. Vastasin näin, koska:
    i. 5 000 kirjaa ei ole kovin paljon ja ne mahtuvat reppuun
    ii. 5 000 g on liian suuri paino nostettavaksi ja suuhun ei voi mahtua 5 000 M&M -karkkia. Siksi oikea vastaus on A.
    iii. Arvasin
    iv. Muu

B. Vastasin näin, koska:
    i. 5 000 g on 5 kg ja sellainen paino on mahdollista nostaa
    ii. Reppuun ei mahdu 5 000 koulukirjaa, koska se on liian suuri määrä ja suuhun ei voi mahtua 5 000 M&M -karkkia. Siksi oikea vastaus on B.
    iii. Arvasin
    iv. Muu

C. Vastasin näin, koska:
    i. Jos Sofi pitää M&M -karkeista, hän voi laittaa niitä suuhunsa kerralla 5 000 kpl
    ii. Reppuun ei mahdu 5 000 koulukirjaa, koska se on liian suuri määrä ja 5 000 g on liian suuri paino nostettavaksi. Siksi oikea vastaus on C.

iii. Arvasin

iv. Muu

D. Vastasin näin, koska:

i. Tarvitsen lisää tietoa, jotta voin vastata

ii. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata

iii. Arvasin

iv. Muu

2. Kun lasketaan yhteen kaksi lukua, joissa on kummassakin kolme numeroa, kuinka monta numeroa on niiden summassa?

a. 3 numeroa

b. 4 numeroa

c. 3 tai 4 numeroa

d. 6 numeroa

e. En osaa sanoa

A. Vastasin näin, koska:

i. Jos kummassakin luvussa on kolme numeroa, niiden summassa on myös kolme numeroa

ii. Jos lasken esimerkiksi 100 + 100 = 200, luvussa 200 on kolme numeroa.

iii. Arvasin

iv. Muu

B. Vastasin näin, koska:

i. Yhteenlaskussa luvut muuttuvat isommiksi, siksi summassa täytyy olla neljä numeroa.

ii. Jos lasken esimerkiksi 500 + 500 = 1000, luvussa 1000 on neljä numeroa.

iii. Arvasin

iv. Muu

C. Vastasin näin, koska:

i. Jos yhteenlaskettavat luvut ovat pieniä, summa on pienempi kuin 1 000, jolloin siinä on kolme numeroa. Jos taas yhteenlaskettavat luvut ovat suurempia, summa on suurempi kuin 1 000, jolloin siinä neljä numeroa.

ii. Jos lasken 100 + 100 = 200, luvussa 200 on kolme numeroa, mutta jos lasken 500 + 500 = 1 000, luvussa 1 000 on neljä numeroa.

iii. Arvasin

iv. Muu

D. Vastasin näin, koska:

i. Kummakin yhteenlaskettavassa luvussa on kolme numeroa ja 3 + 3 = 6, summassa on kuusi numeroa.

ii. Arvasin

iii. Muu

E. Vastasin näin, koska:

i. En muista, mitä ”luku” tai ”numero” tarkoittavat.

ii. Minun pitäisi tietää, mitkä yhteenlaskettavat luvut ovat, jotta voisin vastata

iii. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata

iv. Arvasin

v. Muu

3. 10 -vuotias Emil on 130 cm pitkä. Hänen pituutensa 20 -vuotiaana voisi olla:

   a. 170 cm

   b. 260 cm

   c. 350 cm

   d. En osaa sanoa

A. Vastasin näin, koska:

   i. Ihmisen on mahdollista olla 170 cm pitkä

   ii. Pituudet 260 cm ja 350 cm ovat liian suuria, siksi oikea vaihtoehto on 170 cm.

   iii. Arvasin

   iv. Muu

B. Vastasin näin, koska:

   i. Emil on kaksi kertaa vanhempi (2 x 10 = 20), hän on myös kaksi kertaa pidempi (2 x 130 cm = 260 cm)

   ii. Pituus 170 cm on liian pieni ja 350 cm on liian suuri, siksi oikea vaihtoehto on 260 cm.

   iii. Arvasin

   iv. Muu

C. Vastasin näin, koska:

   i. Ihmisen on mahdollista olla 350 cm pitkä.

ii. Pituudet 170 cm ja 260 cm ovat liian pieniä, siksi oikea vaihtoehto on 350 cm.

iii. Arvasin

iv. Muu

D. Vastasin näin, koska:

i. Ei voi ennakoida ihmisen pituutta.

ii. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata

iii. Arvasin

iv. Muu

4. Minna sanoo ”Voin juoda 5 _____ maitoa joka aamu.” Mikä seuraavista suureista voisi sopia täydentämään lauseen:

a. dl

b. l

c. En osaa sanoa

A. Vastasin näin, koska:

i. 5 dl on puoli litraa eli puoli purkkia maitoa. On mahdollista juoda sen verran maitoa päivässä.

ii. 5 l on liian paljon. Oikea vaihtoehto on siis A.

iii. Arvasin

iv. Muu

B. Vastasin näin, koska:

i. Maitoa mitataan yleensä litroissa, joten oikea vaihtoehto on B.

ii. dl on liian pieni yksikkö, joten oikea vaihtoehto on B.

iii. Arvasin

iv. Muu

C. Vastasin näin, koska:

i. Jokainen juo eri verran maitoa, eli vastaus riippuu henkilöstä.

ii. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata

iii. Arvasin

iv. Muu

5. Luokkahuoneen korkeus voisi olla:

a. 300 cm

b. 300 m

c. En osaa sanoa

A. Vastasin näin, koska:

i. 300 cm on 3 m, ja huoneen korkeus voi olla 3 m.

ii. 300 m on liian suuri etäisyys, joten 300 cm on oikea vastaus.

iii. Arvasin

iv. Muu

B. Vastasin näin, koska:

i. Huoneen korkeutta mitataan yleensä metreissä, joten oikea vastaus on 300 m.

ii. 300 cm on liian pieni etäisyys, joten 300 m on oikea vastaus.

iii. Arvasin

iv. Muu

C. Vastasin näin, koska:

i. Minun pitäisi nähdä luokkahuone, jotta voisin vastata.

ii. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata.

iii. Arvasin

iv. Muu

6. Mikä seuraavista vaihtoehdoista on paras vastaus laskutoimitukselle 149 x 4= ___:

a. alle 500

b. 500:n ja 600:n välissä

c. 600:n ja 700:n välissä

d. En osaa sanoa

A. Vastasin näin, koska:

i. 149 on lähellä 100:a ja 100 x 4 = 400, mikä on alle 500.

ii. 600 ja 700 ovat liian isoja lukuja, joten vastaus on ”alle 500”.

iii. Arvasin

iv. Muu

B. Vastasin näin, koska:

i. 149 on vähän alle 150, ja 150 x 4 = 600, joten tulo 149 x 4 on vähän alle 600.

ii. Laskin 149 x 4 = 596, mikä on 500:n ja 600:n välissä.

iii. Arvasin

iv. Muu

C. Vastasin näin, koska:

    i. 149 on lähellä 150:ä, ja 150 x 4 = 600, joten tulo 149 x 4 on 600:n ja 700:n välissä

    ii. Arvasin

    iii. Muu

D. Vastasin näin, koska:

    i. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata.

    ii. Arvasin

    iii. Muu

7. Minkä luvun sijoittaisit laatikkoon (458 - ____), jotta vastaus on mahdollisimman suuri?

    a. 202

    b. 210

    c. 258

    d. En osaa sanoa

A. Vastasin näin, koska:

    i. Jotta saan suuren vastauksen minun pitää vähentää pienin luku ja 202 on pienempi kuin 210 ja 258.

    ii. Tein kaikki laskutoimitukset. Sain suurimman vastauksen, kun vähensin luvun 202.

    iii. Arvasin

    iv. Muu

B. Vastasin näin, koska:

i. Tein kaikki laskutoimitukset. Sain suurimman vastauksen, kun vähensin luvun 210.

ii. Arvasin

iii. Muu

C. Vastasin näin, koska:

i. 258 on isompi kuin 202 ja 210, joten saan suuren vastauksen vähentämällä 258.

ii. Tein kaikki laskutoimitukset. Sain suurimman vastauksen, kun vähensin luvun 258.

iii. Arvasin

iv. Muu

D. Vastasin näin, koska:

i. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata.

ii. Arvasin

iii. Muu

8. 1,5 kg ruokaa A maksaa 50 euroa ja 3,5 kg ruokaa B maksaa 100 euroa. Kumpi ruoka on halvempaa?

a. ruoka A

b. ruoka B

c. kumpikin ovat yhtä halpoja

d. En osaa sanoa

A. Vastasin näin, koska:

i. 50 euroa on vähemmän kuin 100 euroa, joten ruoka A on halvempaa.

ii. Jaoin 50 euroa 1,5:llä ja 100 euroa 3,5:llä, ja sain siten kummankin ruoan kilohinnan. Sain selville, että ruoka A on halvempaa.

iii. Arvasin

iv. Muu

B. Vastasin näin, koska:

i. Jaoin 50 euroa 1,5:llä ja 100 euroa 3,5:llä, ja sain siten kummankin ruoan kilohinnan. Sain selville, että ruoka B on halvempaa.

ii. 100 eurolla voin ostaa 3 kg ruokaa A tai 3,5 kg ruokaa B, joten ruoka B on halvempaa.

iii. Arvasin

iv. Muu

C. Vastasin näin, koska:

i. Jaoin 50 euroa 1,5:llä ja 100 euroa 3,5:llä, ja sain selville, että ruoat A ja B ovat yhtä halpoja.

ii. Arvasin

iii. Muu

D. Vastasin näin, koska:

i. Minun pitäisi tietää ruokien kilohinta, jotta voisin verrata niitä toisiinsa.

ii. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata.

iii. Arvasin

iv. Muu

9. Mikä seuraavista vaihtoehdoista on paras vastaus laskutoimitukselle 37/49 + 12/49 = ____?

a. alle 1

b. 1

c. enemmän kuin 1

d. en osaa sanoa

A. Vastasin näin, koska:

    i. Kahden murtoluvun summa on murtoluku ja murtoluvut ovat pienempiä kuin 1.

    ii. Laskin 37/49 + 12/49 ja vastaus on pienempi kuin 1.

    iii. Arvasin

    iv. Muu

B. Vastasin näin, koska:

    i. 37/49 on pienempi kuin 1 ja 12/49 on pienempi kuin 1, joten niiden summan täytyy olla 1.

    ii. Laskin 37 + 12 = 49 ja 49/49 on yhtä suuri kuin 1.

    iii. Arvasin

    iv. Muu

C. Vastasin näin, koska:

    i. 37 ja 12 ovat kummatkin isompia kuin 1, joten niiden summa on myös isompi kuin 1.

    ii. Laskin 37 + 12 = 49 ja 49 on isompi kuin 1.

    iii. Arvasin

    iv. Muu

D. Vastasin näin, koska:

    i. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata.

ii. Arvasin

iii. Muu

10. Kun lasketaan yhteen kaksi lukua, joissa on kummassakin yksi desimaali, kuinka monta desimaalia on niiden summassa?

a. 0 desimaalia

b. 1 desimaali

c. 0 tai 1 desimaalia

d. 2 desimaalia

e. En osaa sanoa

A. Vastasin näin, koska:

i. Jos lasken esimerkiksi 0,5 + 0,5 = 1, ja luvussa 1 on 0 desimaalia.

ii. Arvasin

iii. Muu

B. Vastasin näin, koska:

i. Jos kummassakin luvussa on yksi desimaali, niiden summassa on myös yksi desimaali.

ii. Jos lasken esimerkiksi 0,1 + 0,1 = 0,2, ja luvussa 0,2 on yksi desimaali.

iii. Arvasin

iv. Muu

C. Vastasin näin, koska:

i. Jos lasken esimerkiksi 0,5 + 0,5 = 1, luvussa 1 on 0 desimaalia. Mutta jos lasken esimerkiksi 0,1 + 0,1 = 0,2, luvussa 0,2 on yksi desimaali.

ii. Jos desimaalien summa on 10, vastauksessa on 0 desimaalia. Muuten vastauksessa on yksi desimaali.

iii. Arvasin

iv. Muu

D. Vastasin näin, koska:

i. Jos kummassakin luvussa on yksi desimaali, niiden summassa on kaksi desimaalia , koska 1 + 1 = 2.

ii. Jos lasken esimerkiksi 0,5 + 0,6 = 0,11, ja luvussa 0,11 on 2 desimaalia.

iii. Arvasin

iv. Muu

E. Vastasin näin, koska:

i. En muista, mitä ”desimaali” tarkoittaa.

ii. Minun pitäisi tietää, mitkä yhteenlaskettavat luvut ovat, jotta voisin vastata

iii. Minun pitäisi tehdä laskutoimituksia paperilla, jotta voin vastata

iv. Arvasin

v. Muu